GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:23:54 ; Search time 738.572 Seconds
(without alignments)
16735.564 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 2088
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2087.6	100.0	2088	3	AAZ87583	Aaz87583 Prostate
	2	2076	99.4	2087	4	AAS04000	Aas04000 Biomarker
	3	1961.6	93.9	2506	3	AAZ87584	Aaz87584 Prostate
	4	1950	93.4	2505	4	AAS04001	Aas04001 Biomarker
	5	275.4	13.2	757	2	AAV16883	Aav16883 Human pro
	6	275.4	13.2	757	2	AAX26018	Aax26018 Prostate
	7	275.4	13.2	757	3	AAZ87503	Aaz87503 Prostate,
	8	275.4	13.2	757	4	AAS03722	Aas03722 Biomarker
c	9	115.6	5.5	573	3	AAA16415	Aaa16415 Human col
	10	115.6	5.5	4051	5	AAS79336	Aas79336 DNA encod
c	11	115.2	5.5	430	6	ABT07086	Abt07086 Human ova
c	12	115.2	5.5	430	8	ABX72964	Abx72964 Human ova
	13	115.2	5.5	1167	5	AAS71702	Aas71702 DNA encod
c	14	115.2	5.5	1167	5	AAS92271	Aas92271 DNA encod
c	15	115.2	5.5	1167	5	AAS75366	Aas75366 DNA encod
c	16	115.2	5.5	1167	5	AAS79702	Aas79702 DNA encod
	17	115.2	5.5	1167	5	AAS84973	Aas84973 DNA encod
	18	115.2	5.5	1205	5	AAS81380	Aas81380 DNA encod
	19	115.2	5.5	1205	5	AAS72059	Aas72059 DNA encod
	20	115.2	5.5	1205	5	AAS71116	Aas71116 DNA encod
	21	115.2	5.5	1364	5	AAS71114	Aas71114 DNA encod
	22	115.2	5.5	1364	5	AAS72053	Aas72053 DNA encod
	23	115.2	5.5	1548	5	AAS71240	Aas71240 DNA encod
	24	115.2	5.5	1719	5	AAS70932	Aas70932 DNA encod
	25	115.2	5.5	1719	5	AAS91719	Aas91719 DNA encod
	26	115.2	5.5	1719	5	AAS69594	Aas69594 DNA encod
c	27	115.2	5.5	1719	5	AAS73101	Aas73101 DNA encod
	28	115.2	5.5	1719	5	AAS83831	Aas83831 DNA encod
	29	115.2	5.5	2160	5	AAS64322	Aas64322 DNA encod
	30	115.2	5.5	2341	5	AAS85521	Aas85521 DNA encod
	31	115.2	5.5	2385	5	AAS81821	Aas81821 DNA encod
c	32	115.2	5.5	2457	5	AAS73106	Aas73106 DNA encod
c	33	115.2	5.5	2523	5	AAS74903	Aas74903 DNA encod
	34	115.2	5.5	2580	5	AAS78969	Aas78969 DNA encod
	35	115.2	5.5	2648	5	AAS71603	Aas71603 DNA encod
	36	115.2	5.5	2685	5	AAS73165	Aas73165 DNA encod
	37	115.2	5.5	2686	5	AAS77246	Aas77246 DNA encod
c	38	115.2	5.5	2748	5	AAS71650	Aas71650 DNA encod
	39	115.2	5.5	2748	5	AAS76929	Aas76929 DNA encod
	40	115.2	5.5	2748	5	AAS72643	Aas72643 DNA encod
	41	115.2	5.5	2748	5	AAS76709	Aas76709 DNA encod
	42	115.2	5.5	2748	5	AAS66493	Aas66493 DNA encod
	43	115.2	5.5	2748	5	AAS78788	Aas78788 DNA encod
	44	115.2	5.5	2955	5	AAS71125	Aas71125 DNA encod
c	45	115.2	5.5	3048	5	AAS92377	Aas92377 DNA encod

ALIGNMENTS

RESULT 1
AAZ87583
ID AAZ87583 standard; DNA; 2088 BP.
XX
AC AAZ87583;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59295.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Claim 1; Page 182-183; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid

CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;

Query Match 100.0%; Score 2087.6; DB 3; Length 2088;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2088; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980

Qy       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040

Qy       2041 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
                 ||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088


RESULT 2
AAS04000
ID   AAS04000 standard; cDNA; 2087 BP.
XX
AC   AAS04000;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             99. .506
FT                   /*tag=  a
FT                   /product= "Prostate cancer marker protein"
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   P-PSDB; AAU02174.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Claim 2; Col 117-121; 78pp; English.
XX
CC   The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC   detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC   sequences can be used as hybridisation probes and primers that
CC   specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC   (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC   nucleic acid markers can be used to produce antibodies for the detection
CC   of prostate, breast or bladder cancer. The nucleic acids can be used as
CC   targets for therapeutic intervention in these diseases, in the
CC   identification and isolation of full-length gene sequences, including
CC   regulatory elements for gene expression, from genomic human DNA
CC   libraries, as hybridisation probes for screening genomic human DNA
CC   libraries. The kits comprising the nucleic acid sequences are useful for
CC   detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ   Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;

  Query Match             99.4%;  Score 2076;  DB 4;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 0;

  Matches 2087;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
```

```
Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1979

Qy       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1980 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2039

Qy       2041 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       2040 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2087
```

RESULT 3
AAZ87584
ID AAZ87584 standard; DNA; 2506 BP.

XX
AC AAZ87584;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 splice variant.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59296.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Claim 1; Page 184-186; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Query Match 93.9%; Score 1961.6; DB 3; Length 2506;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
```

```
Qy       241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
Db       241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy       301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
Db       301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy       361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
Db       361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy       421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
Db       421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy       481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
Db       481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy       541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
Db       541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy       601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
Db       601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy       661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
Db       661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy       721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
Db       721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy       781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
Db       781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
Db      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
Db      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
Db      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
Db      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
Db      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
Db      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
Db      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
Db      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
Db      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy      1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
Db      1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy      1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
Db      1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy      1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
Db      1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy      1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
Db      1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy      1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
Db      1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978
```

```
RESULT 4
AAS04001
ID   AAS04001 standard; cDNA; 2505 BP.
XX
AC   AAS04001;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             99. .506
FT                   /*tag=  a
FT                   /product= "Prostate cancer marker protein"
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
```

XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri R;
XX
DR WPI; 2001-289849/30.
DR P-PSDB; AAU02175.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS Claim 2; Col 121-125; 78pp; English.
XX
CC The sequence represents nucleic acid biomarker UC band 28 #3, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;

Query Match 93.4%; Score 1950; DB 4; Length 2505;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| 
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679
```

```
Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             ||||||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1977
```

RESULT 5
AAV16883
ID AAV16883 standard; DNA; 757 BP.
XX
AC AAV16883;
XX
DT 07-AUG-1998 (first entry)
XX
DE Human prostate cancer marker UC Band #28.
XX
KW Prostate cancer; human; marker; diagnosis; treatment; probe; ss.
XX
OS Homo sapiens.
XX
PN WO9804689-A1.
XX
PD 05-FEB-1998.
XX
PF 31-JUL-1996; 96WO-US012516.
XX
PR 31-JUL-1996; 96WO-US012516.
XX
PA (UROC-) UROCOR INC.
XX
PI Veltri R, Ohara SM, An G, Ralph D;
XX
DR WPI; 1998-130681/12.
XX
PT Human prostate cancer marker - useful for detection and treatment of
PT human prostate cancer.
XX
PS Claim 1; Page 133-134; 229pp; English.
XX
CC This represents a marker sequence for human prostate cancer. Isolated
CC nucleic acid segments shown in AAV16881 to AAV16885, AAV16890 to
CC AAV16903, AAV26351 and AAV26352 which can act as human prostate cancer
CC markers are provided in the specification. It also provides methods for
CC identifying markers for human prostate cancer and for detection of
CC prostate cancer cells. The markers can be identified by amplifying human
CC prostate RNA to provide nucleic acid amplification products, separating
CC the products and identifying those RNA that are differentially expressed
CC between human prostate cancers versus normal or benign human prostate.
CC Prostate cancer cells in a sample can be detected by detecting a nucleic
CC acid in a sample, the nucleic acid being a prostate cancer marker.
CC Primers and probes derived from this marker can be used for the detection
CC of prostate cancer cells in a sample. Antibodies against the protein
CC encoded by the marker nucleic acid fragments, inhibitors of the protein
CC and oligonucleotides antisense to the markers can be used in the
CC treatment of prostate cancer. The antibodies can also be used for the
CC diagnosis of human prostate cancer
XX
SQ Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

```
  Query Match             13.2%;  Score 275.4;  DB 2;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 9.3e-53;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             ||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293
```

RESULT 6
AAX26018
ID AAX26018 standard; DNA; 757 BP.
XX
AC AAX26018;
XX
DT 20-MAY-1999 (first entry)
XX
DE Prostate disease marker gene fragment UC Band #28.
XX
KW Prostate cancer; benign prostatic hyperplasia; marker gene; tumour;
KW differentiation; Reverse Transcription Polymerase Chain Reaction;
KW diagnostic; progression; cancer; metastasis; human; RT-PCR; ss.
XX
OS Homo sapiens.
XX
PN US5882864-A.
XX
PD 16-MAR-1999.
XX
PF 31-JUL-1996; 96US-00692787.
XX
PR 31-JUL-1995; 95US-0001655P.
XX
PA (UROC-) UROCOR INC.
XX
PI Veltri R, Ralph D, An G, O'hara SM;
XX
DR WPI; 1999-214055/18.
XX
PT Diagnosing prostate cancer and benign prostatic hyperplasia cells - using
PT oligonucleotide probes specific for marker genes associated with tumor
PT differentiation and progression in Reverse Transcription Polymerase Chain
PT Reaction analysis.
XX
PS Claim 1; Col 71-72; 74pp; English.
XX
CC The invention relates to methods for diagnosing prostate cancer or benign
CC prostatic hyperplasia cells in a biological sample. The method uses
CC oligonucleotide probes specific for marker genes associated with tumour
CC differentiation and progression in Reverse Transcription Polymerase Chain
CC Reaction (RT-PCR) analysis. The methods are diagnostic techniques useful
CC for detecting and monitoring the progression of benign prostatic
CC hyperplasia and human prostate cancer (the most prevalent form of cancer
CC and a major cause of death in males) prior to the tumor undergoing
CC metastasis, therefore allowing the optimal method of treatment to be
CC determined before the condition becomes life threatening. The present

```
CC   sequence represents a claimed marker gene fragment
XX
SQ   Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

  Query Match             13.2%;  Score 275.4;  DB 2;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 9.3e-53;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293


RESULT 7
AAZ87503
ID   AAZ87503 standard; cDNA; 757 BP.
XX
AC   AAZ87503;
XX
DT   19-APR-2000  (first entry)
XX
DE   Prostate, breast and bladder cancers detecting biomarker UC Band #28.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW   diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Claim 1; Page 165; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

  Query Match             13.2%;  Score 275.4;  DB 3;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 9.3e-53;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293


RESULT 8
AAS03722
ID   AAS03722 standard; cDNA; 757 BP.
XX
AC   AAS03722;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28 #1, used in diagnosis and prognosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Claim 2; Col 79; 78pp; English.
XX
```

```
CC   The sequence represents nucleic acid biomarker UC band 28 #1, used in
CC   detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC   sequences can be used as hybridisation probes and primers that
CC   specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC   (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC   nucleic acid markers can be used to produce antibodies for the detection
CC   of prostate, breast or bladder cancer. The nucleic acids can be used as
CC   targets for therapeutic intervention in these diseases, in the
CC   identification and isolation of full-length gene sequences, including
CC   regulatory elements for gene expression, from genomic human DNA
CC   libraries, as hybridisation probes for screening genomic human DNA
CC   libraries. The kits comprising the nucleic acid sequences are useful for
CC   detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ   Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

  Query Match             13.2%;  Score 275.4;  DB 4;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 9.3e-53;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db             1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||    |  |||  ||  | |
Db           241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293


RESULT 9
AAA16415/c
ID   AAA16415 standard; DNA; 573 BP.
XX
AC   AAA16415;
XX
DT   14-JUN-2000  (first entry)
XX
DE   Human colon cancer differentially expressed nucleotide sequence #420.
XX
KW   Colon cancer; detect; differential expression; human; treatment;
KW   detect mutation; non-invasive diagnostic method; ds.
XX
OS   Homo sapiens.
XX
PN   WO200012702-A2.
XX
PD   09-MAR-2000.
XX
PF   30-AUG-1999;   99WO-US019424.
XX
PR   31-AUG-1998;   98US-0098639P.
PR   27-JAN-1999;   99US-0117393P.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Endege WO,  Steinmann KE,  Astle JH,  Burgess CC,  Carroll E;
PI   Catino TJ,  Dwivedi P,  Ford DM,  Lewis ME,  Molino GA,  Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-256641/22.
XX
PT   Novel nucleic acids and proteins for identifying therapeutic agents
PT   useful for treating and diagnosing cancer, especially colon cancer.
XX
PS   Claim 16; Page 294; 345pp; English.
XX
CC   This sequence represents a human nucleotide sequence which is
CC   differentially expressed in colon cancer cells compared to the expression
CC   levels in normal cells. The nucleotide sequence can be used as a source
CC   of primers and probes. The nucleotide sequence is useful for determining
CC   the phenotype of a cell by detecting the differential expression of the
CC   sequence relative to a normal cell. The probes derived from the sequence
CC   can also be used to determine the phenotype of cells in a sample. Probes
CC   and antibodies which hybridise to the nucleotide sequence can also be
CC   used to determine the phenotype of a cell. The primers are useful for
CC   detecting a mutation in a test nucleotide sequence and also for detecting
CC   cancer, preferably colon cancer. Antibodies against the protein encoded
CC   by the nucleotide sequence can also be used in a method to detect colon
CC   cancer. The diagnostic method is non-invasive and accurate for diagnosing
CC   colon cancer at an early stage
XX
SQ   Sequence 573 BP; 98 A; 91 C; 111 G; 236 T; 0 U; 37 Other;

  Query Match              5.5%;  Score 115.6;  DB 3;  Length 573;
  Best Local Similarity   95.2%;  Pred. No. 2.4e-16;
  Matches  118;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1965 CGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGG 2024
             | | |  | ||||||||||||||||||||||||||| ||||||||||||||||||||
Db            312 CAAGACTANATAAAAAAAGAGAGAAGAATCAAATAGACNCAATAAAAAATGATAAAGGGG 253

Qy       2025 ATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCT 2084
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            252 ATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCT 193

Qy       2085 ACGC 2088
             ||||
Db            192 ACGC 189


RESULT 10
AAS79336
ID   AAS79336 standard; cDNA; 4051 BP.
XX
AC   AAS79336;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #15140.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG15149.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
```

```
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 15140; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4051 BP; 1562 A; 824 C; 819 G; 846 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 115.6;  DB 5;  Length 4051;
  Best Local Similarity   89.9%;  Pred. No. 4.3e-16;
  Matches  124;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1951 AGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAA 2010
              | ||||     | |   |||| |  | ||||||||||||||||||||||||||||||
Db       1807 ATACTCCTTGAGAAGAGCAACTCCAAGAAAAAGAGAGAAGAATCAAATAGACACAATAAA 1866

Qy       2011 AAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATA 2070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1867 AAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATA 1926

Qy       2071 CTACAAACACCTCTACGC 2088
              ||||||||||||||||||
Db       1927 CTACAAACACCTCTACGC 1944


RESULT 11
ABT07086/c
ID   ABT07086 standard; cDNA; 430 BP.
XX
AC   ABT07086;
XX
DT   07-NOV-2002  (first entry)
XX
DE   Human ovarian cancer associated coding sequence SEQ ID NO: 248.
XX
KW   Human; ovarian cancer; cancer; gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2002076715-A1.
XX
PD   20-JUN-2002.
XX
PF   06-JUN-2001; 2001US-00876889.
XX
PR   23-SEP-1998;   98US-00159320.
PR   08-FEB-1999;   99US-00246429.
PR   16-SEP-1999;   99US-00397787.
XX
PA   (BENS/) BENSON D R.
PA   (LODE/) LODES M J.
PA   (MITC/) MITCHAM J L.
PA   (KING/) KING G E.
XX
PI   Benson DR,  Lodes MJ,  Mitcham JL,  King GE;
XX
DR   WPI; 2002-598720/64.
XX
PT   Composition for detecting and treating ovarian cancer, comprises a
PT   specific polypeptide, polynucleotide, T cell population, or antigen
PT   presenting cell.
XX
PS   Example 1; Page 94; 188pp; English.
XX
CC   The present invention relates to a method of detecting the presence of
CC   ovarian cancer in a patient, involving detecting ovarian cancer
CC   associated polynucleotides. The method is not only used to detect the
CC   presence of cancer, preferably ovarian cancer in a patient, but also is
CC   used to stimulate and/or expand T cells specific for an ovarian tumour
CC   protein. The sequences can be used in vaccines used to treat cancer. The
CC   present sequence is an ovarian cancer associated coding sequence
XX
SQ   Sequence 430 BP; 70 A; 74 C; 85 G; 196 T; 0 U; 5 Other;

  Query Match             5.5%;  Score 115.2;  DB 6;  Length 430;
  Best Local Similarity   97.5%;  Pred. No. 2.7e-16;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 AATAAAGAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 236

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 176


RESULT 12
ABX72964/c
ID   ABX72964 standard; cDNA; 430 BP.
XX
AC   ABX72964;
XX
DT   14-MAR-2003  (first entry)
XX
DE   Human ovarian carcinoma antigen partial cDNA sequence #230.
XX
KW   Human; cancer detection; ovarian carcinoma antigen; ovarian cancer;
KW   tumour antigen; tumour; OV2; OV3; OV6; OV9; OV10; OV12; OV14; OV17; OV18;
KW   OV23; OV24; OV27; OV41; OV54; OV57; gene; ss.
XX
OS   Homo sapiens.
XX
PN   US6468758-B1.
XX
PD   22-OCT-2002.
XX
PF   16-SEP-1999;   99US-00397787.
XX
PR   23-SEP-1998;   98US-00159320.
PR   08-FEB-1999;   99US-00246429.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Benson DR,  Lodes MJ,  Mitcham JL,  King GE;
XX
DR   WPI; 2003-147101/14.
XX
PT   Determining presence or absence of cancer in patient by contacting
PT   patient sample with oligonucleotide that hybridizes to polynucleotide
PT   encoding ovarian carcinoma antigen, and detecting amount of hybridization
PT   in sample.
XX
PS   Example 1; Col 181-182; 152pp; English.
XX
CC   The present invention relates to compositions and methods for determining
```

```
CC   presence or absence of cancer in patient. The method comprising
CC   contacting a biological sample with an oligonucleotide that hybridises to
CC   partial polynucleotide sequence encoding for human ovarian carcinoma
CC   antigen, or its complement. The method and compositions are useful for
CC   the therapy and diagnosis of ovarian cancer. The method can be used to
CC   identify tumour antigens that are secreted from ovarian carcinoma and/or
CC   other tumours. Effective cancer detection is achieved using the method of
CC   the invention. ABX72720-ABX73050 represent partial cDNA sequences
CC   encoding human ovarian carcinoma antigens
XX
SQ   Sequence 430 BP; 70 A; 74 C; 85 G; 196 T; 0 U; 5 Other;

  Query Match             5.5%;  Score 115.2;  DB 8;  Length 430;
  Best Local Similarity   97.5%;  Pred. No. 2.7e-16;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 AATAAAGAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 236

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 176


RESULT 13
AAS71702
ID   AAS71702 standard; cDNA; 1167 BP.
XX
AC   AAS71702;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #7506.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG07515.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 7506; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1167 BP; 492 A; 264 C; 202 G; 209 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 115.2;  DB 5;  Length 1167;
  Best Local Similarity   97.5%;  Pred. No. 3.7e-16;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 AATAAAGAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 386

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 446


RESULT 14
AAS92271/c
ID   AAS92271 standard; cDNA; 1167 BP.
XX
AC   AAS92271;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28075.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28084.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 28075; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
```

```
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1167 BP; 209 A; 202 C; 264 G; 492 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 115.2;  DB 5;  Length 1167;
  Best Local Similarity   97.5%;  Pred. No. 3.7e-16;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATAAAGAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 782

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 722


RESULT 15
AAS75366/c
ID   AAS75366 standard; cDNA; 1167 BP.
XX
AC   AAS75366;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #11170.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG11179.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 11170; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1167 BP; 209 A; 202 C; 264 G; 492 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 115.2;  DB 5;  Length 1167;
  Best Local Similarity   97.5%;  Pred. No. 3.7e-16;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATAAAGAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 782

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 722


Search completed: September  8, 2005, 23:49:42
Job time : 743.572 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 06:16:18 ; Search time 99.5368 Seconds
(without alignments)
16226.276 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174357_19482/app_query.fasta_1.4942
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_232_@runat_07092005_174357_19482 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	19.3	135	3	AAY59296	Aay59296 Prostate
2	694	19.3	135	3	AAY59295	Aay59295 Prostate
3	694	19.3	135	4	AAU02174	Aau02174 Biomarker
4	694	19.3	135	4	AAU02175	Aau02175 Biomarker
5	206.5	5.8	717	4	ABG06939	Abg06939 Novel hum
6	206.5	5.8	1746	4	ABG05644	Abg05644 Novel hum
7	206.5	5.8	1824	4	ABG09152	Abg09152 Novel hum
8	204	5.7	724	4	ABG15384	Abg15384 Novel hum
9	204	5.7	724	4	ABG17531	Abg17531 Novel hum
10	204	5.7	1257	4	ABG17335	Abg17335 Novel hum
11	204	5.7	1921	4	ABG06576	Abg06576 Novel hum
12	203	5.7	625	4	ABG11195	Abg11195 Novel hum
13	203	5.7	626	4	ABG17324	Abg17324 Novel hum
14	202	5.6	590	4	ABG06851	Abg06851 Novel hum
15	202	5.6	857	4	ABG04403	Abg04403 Novel hum
16	202	5.6	954	4	ABG02322	Abg02322 Novel hum
17	202	5.6	954	4	ABG08451	Abg08451 Novel hum
18	202	5.6	956	4	ABG14785	Abg14785 Novel hum
19	202	5.6	956	4	ABG14541	Abg14541 Novel hum
20	202	5.6	956	5	ABG77047	Abg77047 Prostate
21	201.5	5.6	794	4	ABG17634	Abg17634 Novel hum
22	201.5	5.6	1557	4	ABG02512	Abg02512 Novel hum
23	201.5	5.6	1709	4	ABG03979	Abg03979 Novel hum
24	201.5	5.6	1709	4	ABG06055	Abg06055 Novel hum
25	201.5	5.6	1709	4	ABG19873	Abg19873 Novel hum
26	201.5	5.6	1709	4	ABG03052	Abg03052 Novel hum
27	201.5	5.6	1709	4	ABG10797	Abg10797 Novel hum
28	201.5	5.6	1709	4	ABG06602	Abg06602 Novel hum
29	201.5	5.6	1709	4	ABG20779	Abg20779 Novel hum
30	201.5	5.6	1709	4	ABG08635	Abg08635 Novel hum
31	201.5	5.6	1709	4	ABG02341	Abg02341 Novel hum
32	201.5	5.6	1709	4	ABG09638	Abg09638 Novel hum
33	201.5	5.6	1809	4	ABG01006	Abg01006 Novel hum
34	201	5.6	1129	8	ADP29913	Adp29913 Human sec
35	200	5.6	302	4	ABG14081	Abg14081 Novel hum
36	200	5.6	520	4	ABG03512	Abg03512 Novel hum
37	200	5.6	985	4	ABG06938	Abg06938 Novel hum
38	200	5.6	1065	4	ABG02332	Abg02332 Novel hum
39	200	5.6	1078	4	ABG10803	Abg10803 Novel hum
40	200	5.6	1078	4	ABG06060	Abg06060 Novel hum
41	200	5.6	1078	4	ABG06606	Abg06606 Novel hum
42	200	5.6	1078	4	ABG09642	Abg09642 Novel hum
43	200	5.6	1085	4	ABG14957	Abg14957 Novel hum
44	200	5.6	1196	4	ABG14610	Abg14610 Novel hum
45	200	5.6	1245	7	ADE09432	Ade09432 Novel pro

ALIGNMENTS

RESULT 1
AAY59296
ID AAY59296 standard; peptide; 135 AA.
XX
AC AAY59296;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AAZ87584.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 184-186; 191pp; English.

XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 135 AA;

Alignment Scores:

Pred. No.:	1.14e-75	Length:	135
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	19.34%	Indels:	0
DB:	3	Gaps:	0

US-09-974-546C-83 (1-2088) x AAY59296 (1-135)

```
Qy         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
```

RESULT 2
AAY59295
ID AAY59295 standard; peptide; 135 AA.
XX
AC AAY59295;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.

XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AAZ87583.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 182-183; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 135 AA;

Alignment Scores:

Pred. No.:	1.14e-75	Length:	135
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	19.34%	Indels:	0
DB:	3	Gaps:	0

US-09-974-546C-83 (1-2088) x AAY59295 (1-135)

```
Qy         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 11:23:58 ; Search time 21.589 Seconds
(without alignments)
14439.486 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174359_19522/app_query.fasta_1.4942
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_42_@runat_07092005_174359_19522 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	771	21.5	157	4	US-09-949-016-9182	Sequence 9182, Ap
2	694	19.3	135	3	US-09-097-199-84	Sequence 84, Appl
3	694	19.3	135	3	US-09-097-199-86	Sequence 86, Appl
4	117.5	3.3	57	4	US-09-513-999C-7461	Sequence 7461, Ap
5	103	2.9	21	3	US-09-097-199-56	Sequence 56, Appl
6	100	2.8	3169	2	US-08-477-451-6	Sequence 6, Appli
7	91.5	2.5	944	3	US-09-134-001C-4352	Sequence 4352, Ap
8	90.5	2.5	1720	2	US-08-477-451-12	Sequence 12, Appl
9	89	2.5	2184	4	US-09-417-485D-6	Sequence 6, Appli
10	88	2.5	305	4	US-09-328-352-6445	Sequence 6445, Ap
11	87	2.4	2210	4	US-09-309-572-7	Sequence 7, Appli
12	87	2.4	2210	4	US-09-718-096-7	Sequence 7, Appli
13	85	2.4	192	4	US-09-270-767-34244	Sequence 34244, A
14	85	2.4	192	4	US-09-270-767-49461	Sequence 49461, A
c 15	84.5	2.4	944	3	US-09-134-001C-4352	Sequence 4352, Ap
16	83	2.3	698	4	US-09-538-092-151	Sequence 151, App
c 17	82	2.3	318	4	US-09-393-634-47	Sequence 47, Appl
c 18	82	2.3	318	4	US-09-949-016-9253	Sequence 9253, Ap
19	81.5	2.3	944	4	US-09-107-532A-4864	Sequence 4864, Ap
20	81	2.3	128	6	5514582-34	Patent No. 5514582
21	81	2.3	128	6	5514582-34	Patent No. 5514582
22	80	2.2	979	4	US-09-538-092-482	Sequence 482, App
23	79.5	2.2	373	4	US-09-270-767-62380	Sequence 62380, A
24	79.5	2.2	671	4	US-09-270-767-46759	Sequence 46759, A
25	79.5	2.2	1151	4	US-09-328-352-5876	Sequence 5876, Ap
26	78.5	2.2	336	3	US-08-804-166-8	Sequence 8, Appli
27	78.5	2.2	336	3	US-08-910-991-8	Sequence 8, Appli
28	78.5	2.2	336	4	US-09-756-186-8	Sequence 8, Appli
29	78.5	2.2	998	4	US-09-949-016-8326	Sequence 8326, Ap
30	78	2.2	180	4	US-09-134-000C-6251	Sequence 6251, Ap
31	77.5	2.2	251	4	US-09-248-796A-24724	Sequence 24724, A
32	77	2.1	441	4	US-09-134-000C-6563	Sequence 6563, Ap
33	77	2.1	515	4	US-09-635-872A-6	Sequence 6, Appli
34	77	2.1	515	4	US-09-636-077A-6	Sequence 6, Appli
35	77	2.1	515	4	US-09-636-060C-6	Sequence 6, Appli
36	77	2.1	515	4	US-09-986-552-6	Sequence 6, Appli
37	77	2.1	515	4	US-09-636-596C-6	Sequence 6, Appli
38	77	2.1	515	4	US-10-023-894-18	Sequence 18, Appl
39	77	2.1	515	4	US-10-306-686-6	Sequence 6, Appli
40	77	2.1	666	1	US-08-318-831-3	Sequence 3, Appli
41	77	2.1	814	1	US-08-318-831-2	Sequence 2, Appli
42	76.5	2.1	239	4	US-09-248-796A-25501	Sequence 25501, A
43	76.5	2.1	369	4	US-09-248-796A-18582	Sequence 18582, A
c 44	76	2.1	125	4	US-09-270-767-59144	Sequence 59144, A
45	76	2.1	166	4	US-09-252-991A-18131	Sequence 18131, A

ALIGNMENTS

RESULT 1
US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9182
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9182

Alignment Scores:

Pred. No.:	6.22e-85	Length:	157
Score:	771.00	Matches:	155
Percent Similarity:	97.48%	Conservative:	0
Best Local Similarity:	97.48%	Mismatches:	2
Query Match:	21.48%	Indels:	2
DB:	4	Gaps:	0

US-09-974-546C-83 (1-2088) x US-09-949-016-9182 (1-157)

Qy 31 TGTATAATAATTTACAAAATTATTCTTCTATTGCTACAGAGCTACAATTCAATTTACAGT 90

```
                ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 CysIleIleIleTyrLysIleIleLeuLeu-CysTyrArgAlaThrIleGlnPheThrVa 20

Qy         91 AGGCCACCATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTC 150
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 l-ValAsnMetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleH 40

Qy        151 ACATTTTACAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCA 210
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         40 isIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyI 60

Qy        211 TTCTAGCTCCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAA 270
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 leLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluL 80

Qy        271 AAGTAAAAAGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACC 330
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 ysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisH 100

Qy        331 ATGCAATTCTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTT 390
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 isAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysS 120

Qy        391 CCTCATTATCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAG 450
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 erSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuG 140

Qy        451 GTCTGGAATGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 lyLeuGluCysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 157
```

RESULT 2
US-09-097-199-84
; Sequence 84, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-097-199-84

Alignment Scores:

Pred. No.:	1.44e-75	Length:	135
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	19.34%	Indels:	0
DB:	3	Gaps:	0

US-09-974-546C-83 (1-2088) x US-09-097-199-84 (1-135)

```
Qy         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
              ||||||||||||||||||||||||||||||||||||||||||||
Db        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
```

RESULT 3
US-09-097-199-86
; Sequence 86, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 12:21:48 ; Search time 117.944 Seconds
(without alignments)
13965.459 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 3554922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174400_19612/app_query.fasta_1.4942
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09974546_@CGN_1_1_229_@runat_07092005_174400_19612
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	19.3	135	10	US-09-974-546-84	Sequence 84, Appl
2	694	19.3	135	10	US-09-974-546-86	Sequence 86, Appl
3	202	5.6	956	14	US-10-000-256A-153	Sequence 153, App
4	197	5.5	917	13	US-10-001-835-189	Sequence 189, App
5	197	5.5	1031	14	US-10-082-828A-257	Sequence 257, App
6	197	5.5	1031	20	US-11-057-807-257	Sequence 257, App
7	197	5.5	1275	14	US-10-025-201-3	Sequence 3, Appli
8	197	5.5	1275	17	US-10-800-322-34	Sequence 34, Appl
9	195	5.4	1397	17	US-10-473-451-6	Sequence 6, Appli
10	192	5.3	1010	14	US-10-050-882-80	Sequence 80, Appl
11	192	5.3	1272	15	US-10-114-270-30	Sequence 30, Appl
12	192	5.3	1274	16	US-10-415-615-3	Sequence 3, Appli
13	191.5	5.3	1081	15	US-10-085-198-150	Sequence 150, App
14	191	5.3	936	15	US-10-243-552-850	Sequence 850, App
15	191	5.3	1018	15	US-10-085-198-156	Sequence 156, App
16	189	5.3	843	15	US-10-001-885-115	Sequence 115, App
17	189	5.3	843	20	US-11-057-447-115	Sequence 115, App
18	183	5.1	990	15	US-10-085-198-154	Sequence 154, App
19	174	4.8	190	14	US-10-074-475-250	Sequence 250, App
20	172	4.8	4961	14	US-10-114-153-64	Sequence 64, Appl
21	170	4.7	940	14	US-10-082-830-235	Sequence 235, App
22	125	3.5	316	13	US-10-001-870-132	Sequence 132, App
23	122	3.4	100	15	US-10-094-749-2789	Sequence 2789, Ap
24	118	3.3	112	11	US-09-864-408A-6702	Sequence 6702, Ap
25	111.5	3.1	127	15	US-10-094-749-3226	Sequence 3226, Ap
26	106	3.0	1017	15	US-10-085-198-152	Sequence 152, App
27	103.5	2.9	371	15	US-10-369-493-20398	Sequence 20398, A
28	103.5	2.9	381	15	US-10-369-493-1334	Sequence 1334, Ap
29	103	2.9	21	10	US-09-974-546-56	Sequence 56, Appl
30	93.5	2.6	6761	17	US-10-732-923-15035	Sequence 15035, A
31	91.5	2.5	461	16	US-10-748-112-21	Sequence 21, Appl
32	91	2.5	35	14	US-10-105-232-510	Sequence 510, App
33	91	2.5	35	14	US-10-189-437-497	Sequence 497, App
34	91	2.5	39	14	US-10-105-232-508	Sequence 508, App
35	91	2.5	39	14	US-10-189-437-495	Sequence 495, App
36	91	2.5	40	14	US-10-105-232-518	Sequence 518, App
37	91	2.5	40	14	US-10-189-437-505	Sequence 505, App
38	91	2.5	44	14	US-10-105-232-514	Sequence 514, App
39	91	2.5	44	14	US-10-105-232-517	Sequence 517, App
40	91	2.5	44	14	US-10-189-437-501	Sequence 501, App
41	91	2.5	44	14	US-10-189-437-504	Sequence 504, App
42	91	2.5	48	14	US-10-105-232-513	Sequence 513, App
43	91	2.5	48	14	US-10-189-437-500	Sequence 500, App
44	91	2.5	56	14	US-10-105-232-507	Sequence 507, App
45	91	2.5	56	14	US-10-189-437-494	Sequence 494, App

ALIGNMENTS

RESULT 1
US-09-974-546-84
; Sequence 84, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:

```
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/974,546
;              FILING DATE: 10-Oct-2001
;              CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 09/097,199
;              FILING DATE: 1998-06-12
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Nakashima, Richard A.
;              REGISTRATION NUMBER: P-42,023
;              REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (512) 418-3000
;              TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 84:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 135 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

Alignment Scores:
Pred. No.:                7.85e-71        Length:         135
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              19.34%          Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-83 (1-2088) x US-09-974-546-84 (1-135)

Qy          99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy         159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy         219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy         279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy         339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy         399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy         459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
               |||||||||||||||||||||||||||||||||||||||||||
Db         121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2
US-09-974-546-86
; Sequence 86, Application US/09974546
; Publication No. US20030050470A1
;   GENERAL INFORMATION:
;        APPLICANT: An, Gang
;                   O'Hara, S. Mark
;                   Ralph, David
;                   Veltri, Robert
;        TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                            PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/974,546
;             FILING DATE: 10-Oct-2001
;             CLASSIFICATION: Unknown
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/097,199
;             FILING DATE: 1998-06-12
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 86:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 135 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Alignment Scores:
Pred. No.:                7.85e-71        Length:         135
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              19.34%          Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-83 (1-2088) x US-09-974-546-86 (1-135)

Qy          99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy         159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy         219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy         279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy         339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy         399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 09:43:33 ; Search time 25.2251 Seconds
(without alignments)
15928.642 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174358_19508/app_query.fasta_1.4942
-DB=PIR_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_79_@runat_07092005_174358_19508 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	197	5.5	1275	2	B28096	line-1 protein ORF
2	197	5.5	1275	2	I38588	reverse transcript
3	197	5.5	1275	2	S65824	reverse transcript
4	195	5.4	1259	4	GNHUL1	retrovirus-related
5	187	5.2	1280	2	B34087	hypothetical prote
6	171.5	4.8	138	2	I59767	lamin-like protein
7	156	4.3	335	4	A26718	retrovirus-related
8	126.5	3.5	1260	4	GNLRL1	retrovirus-related
9	125	3.5	190	2	B25556	hypothetical prote
10	124	3.5	1281	1	GNMSLL	retrovirus-related
11	118	3.3	685	2	S16783	probable RNA-direc
12	115	3.2	500	2	S16788	probable reverse t
13	114	3.2	556	2	S21347	hypothetical prote
14	103.5	2.9	381	2	F71196	probable hexosyltr
15	94	2.6	82	2	D60396	antigen 7H8/7 - ma
16	92.5	2.6	263	2	A82069	hypothetical prote
17	92	2.6	536	2	T37840	probable vacuolar
18	91.5	2.5	461	1	GQRTT1	tumor necrosis fac
c 19	90	2.5	1136	2	T40355	hypothetical prote
20	89.5	2.5	399	2	D86322	hypothetical prote
c 21	87.5	2.4	367	2	T25821	hypothetical prote
22	87	2.4	379	2	C75006	hypothetical prote
23	87	2.4	2210	1	RRXPLC	genome polyprotein
24	86.5	2.4	451	2	D70045	two-component sens
25	86.5	2.4	459	2	S06607	23S rRNA intron 2
26	85.5	2.4	523	2	D64555	conserved hypothet
27	85.5	2.4	3119	2	T18414	protein g377 - mal
c 28	85	2.4	753	2	H90124	hypothetical prote
29	85	2.4	2539	2	B71619	hypothetical prote
30	84.5	2.4	645	2	F86174	hypothetical prote
31	83	2.3	698	2	S52674	general sporulatio
c 32	82.5	2.3	367	2	T21204	hypothetical prote
33	82.5	2.3	472	2	B90095	hypothetical prote
34	82.5	2.3	599	2	AH2184	hypothetical prote
c 35	82	2.3	196	2	G90115	hypothetical prote
36	82	2.3	1258	2	T14855	reverse transcript
37	81.5	2.3	731	2	A99106	hypothetical prote
38	81.5	2.3	1386	1	RNLVC2	DNA-directed RNA p
39	81	2.3	365	2	B84611	probable ubiquitin
40	81	2.3	833	2	T28385	ORF MSV224 probabl
c 41	80.5	2.2	380	2	T25819	hypothetical prote
42	80.5	2.2	434	2	AG2349	hypothetical prote
43	80.5	2.2	479	2	AD1263	two-component sens
c 44	80.5	2.2	680	2	H90093	guanine nucleotide
45	80.5	2.2	706	2	D90124	hypothetical prote

ALIGNMENTS

RESULT 1
B28096
line-1 protein ORF2 - human
C;Species: Homo sapiens (man)
C;Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 09-Jul-2004
C;Accession: B28096
R;Skowronski, J.; Fanning, T.G.; Singer, M.F.
Mol. Cell. Biol. 8, 1385-1397, 1988
A;Title: Unit-length line-1 transcripts in human teratocarcinoma cells.
A;Reference number: A28096; MUID:88246405; PMID:2454389
A;Accession: B28096
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1275 <SKO>
A;Cross-references: UNIPROT:Q15604; UNIPROT:Q9UN80; UNIPROT:Q12881; UNIPROT:O00363; UNIPROT:Q9Y5K0; UNIPROT:O00366; UNIPROT:Q8TE30; UNIPROT:O00375
C;Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 1.05e-12 Length: 1275
Score: 197.00 Matches: 37
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.49% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-83 (1-2088) x B28096 (1-1275)

```
Qy       1976 AAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACC 2035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 LysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThr 402

Qy       2036 GATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTAC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 AspProThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyr 419
```

RESULT 2
I38588

reverse transcriptase homolog - human retrotransposon L1
N;Alternate names: ORF2 protein
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I38588
R;Holmes, S.E.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H.
Nature Genet. 7, 143-148, 1994
A;Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q p
A;Reference number: I38587; MUID:95004577; PMID:7920631
A;Accession: I38588
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1275 <RES>
A;Cross-references: UNIPROT:Q12881; EMBL:U09116; NID:g483914; PIDN:AAB60345.1; PID:g4839
C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	1.05e-12	Length:	1275
Score:	197.00	Matches:	37
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.49%	Indels:	0
DB:	2	Gaps:	0

US-09-974-546C-83 (1-2088) x I38588 (1-1275)

```
Qy       1976 AAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACC 2035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 LysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThr 402

Qy       2036 GATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTAC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 AspProThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyr 419
```

RESULT 3
S65824
reverse transcriptase homolog - human transposon L1.1
C;Species: Homo sapiens (man)
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S65824
R;Dombroski, B.A.
submitted to the EMBL Data Library, January 1992
A;Description: Isolation of an active human transposable element.
A;Reference number: S65823
A;Accession: S65824
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1275 <DOM>
A;Cross-references: UNIPROT:Q15604; EMBL:M80340; NID:g339767; PIDN:AAA51622.1; PID:g3397
C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	1.05e-12	Length:	1275
Score:	197.00	Matches:	37
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.49%	Indels:	0
DB:	2	Gaps:	0

US-09-974-546C-83 (1-2088) x S65824 (1-1275)

```
Qy       1976 AAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACC 2035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 LysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspLysGlyAspIleThrThr 402

Qy       2036 GATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTAC 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 AspProThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLeuTyr 419
```

RESULT 4
GNHUL1
retrovirus-related reverse transcriptase pseudogene - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 09-Jul-2004
C;Accession: A25313
R;Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.
Nature 321, 625-628, 1986
A;Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence
A;Reference number: A93381; MUID:86230917; PMID:2423883
A;Accession: A25313
A;Status: conceptual translation of pseudogene
A;Molecule type: DNA
A;Residues: 1-1259 <HAT>
A;Cross-references: UNIPROT:P08547
A;Note: this sequence was constructed from an alignment of published and unpublished seq
C;Keywords: reverse transcriptase; pseudogene

Alignment Scores:

Pred. No.:	1.74e-12	Length:	1259
Score:	195.00	Matches:	69
Percent Similarity:	47.52%	Conservative:	27
Best Local Similarity:	34.16%	Mismatches:	77
Query Match:	5.43%	Indels:	29
DB:	4	Gaps:	8

US-09-974-546C-83 (1-2088) x GNHUL1 (1-1259)

```
Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||::::::|||:::   :::   :::   |||         :::|||:::
Db        240 LysLeuThrGlnAsnHisSerThrThrTrp---LysLeuAsnAsnLeuLeuLeuAsnAsp 258

Qy       1561 CCTTATGAGCAGGAAGAG--------------GGAAGGCTGTTTGAGAACCATTTACTT 1605
                 :::         |||                 :::   ||||||   :::
Db        259 TyrTrpValHisAsnGluMetLysAlaGluIleLysLysPhePheGluThrAsnGluAsn 278

Qy       1606 AGCAGAACCACATATTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTG--- 1662
                   |||||||||         :::         ||||||:::   :::|||
Db        279 LysAspThrThrTyrGlnAsnLeuTrpAsp------ThrAlaLysAlaValCysArgGly 296

Qy       1663 CAAACTTGTTRGATCAACCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAA 1722
              :::         :::|||      :::|||      :::         |||
Db        297 LysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIle------Asp 314

Qy       1723 ACATTGGATTTTTCCTTAACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGC 1782
              ||||||      |||         |||:::                        |||   ||
Db        315 ThrLeu---IleSerGlnLeuLysGluLeuGluLysGlnGluGlnThrAsn-SerLysAl 333

Qy       1783 TTCAGGCTAAGAAGAGATTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACT 1842
              ||||      :::||||||                        |||:::         ||
Db        333 aSerArgArgGlnGluIleIle--------------------LysIleArgAlaGluLe 346

Qy       1843 CTCAAACTTCACTAATGATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCC 1902
              |                        |||   |||   |||
Db        346 uLysGluIleGluThrGlnLysThrLeuGlnLysIleAsn-------------------- 359

Qy       1903 GAGTTAAGAACATCAAAACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTG 1962
                 |||   ::::::         |||:::   :::::::::
Db        360 -------GluSerArgSerTrpPhePheGluLysIleAsnLysIleAspArgProLeuAl 377

Qy       1963 CACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGG 2022
                        ||||||||||||||||||||||||||||||||||||||||||:::||
Db        377 aArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspArgGl 397

Qy       2023 GGATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCT 2082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 yAspIleThrThrAspProThrGluIleGlnThrThrIleArgGluTyrTyrLysHisLe 417

Qy       2083 CTAC 2086
              ||||
Db        417 uTyr 418
```

RESULT 5
B34087
hypothetical protein (L1H 3' region) - human
C;Species: Homo sapiens (man)
C;Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 09-Jul-2004

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 06:29:13 ; Search time 124.307 Seconds
(without alignments)
17202.865 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174358_19494/app_query.fasta_1.4942
-DB=UniProt_03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_302_@runat_07092005_174358_19494 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	19.3	135	2	Q9GZY1	Q9gzy1 homo sapien
2	200	5.6	1275	2	O00372	O00372 homo sapien
3	198.5	5.5	1275	2	O00375	O00375 homo sapien
4	197	5.5	111	2	Q9UJZ4	Q9ujz4 homo sapien
5	197	5.5	1192	2	Q7KZ41	Q7kz41 homo sapien
6	197	5.5	1275	2	O00360	O00360 homo sapien
7	197	5.5	1275	2	O00366	O00366 homo sapien
8	197	5.5	1275	2	O00368	O00368 homo sapien
9	197	5.5	1275	2	O00370	O00370 homo sapien
10	197	5.5	1275	2	O00378	O00378 homo sapien
11	197	5.5	1275	2	Q8TE30	Q8te30 homo sapien
12	195	5.4	1259	1	LIN1_HUMAN	P08547 homo sapien
13	193.5	5.4	1275	2	O00362	O00362 homo sapien
14	190	5.3	1275	2	O00363	O00363 homo sapien
15	171.5	4.8	138	2	Q14755	Q14755 homo sapien
16	136	3.8	85	2	Q25910	Q25910 plasmodium
17	136	3.8	232	2	Q6ZNW9	Q6znw9 homo sapien
18	128	3.6	1356	2	Q6QI76	Q6qi76 rattus norv
19	126.5	3.5	1260	1	LIN1_NYCCO	P08548 nycticebus
20	125	3.5	190	2	Q63501	Q63501 rattus norv
21	124	3.5	1252	2	Q9QWY2	Q9qwy2 mus musculu
22	124	3.5	1281	2	O54850	O54850 mus musculu
23	124	3.5	1281	2	O88913	O88913 mus musculu
24	124	3.5	1281	2	O88915	O88915 mus musculu
25	124	3.5	1281	2	Q91Z88	Q91z88 mus musculu
26	124	3.5	1281	2	Q792I9	Q792i9 mus musculu
27	124	3.5	1281	2	Q9QUI2	Q9qui2 mus musculu
28	124	3.5	1281	2	Q9QWY0	Q9qwy0 mus musculu
29	124	3.5	1300	1	POL2_MOUSE	P11369 mus musculu
30	124	3.5	1300	2	O08906	O08906 mus musculu
31	123	3.4	1281	2	O88914	O88914 mus musculu
32	123	3.4	1281	2	Q9QWY3	Q9qwy3 mus musculu
33	121.5	3.4	1575	2	Q6TXJ1	Q6txj1 rattus norv
34	121	3.4	1275	2	O62658	O62658 canis famil
35	120	3.3	1281	2	Q91Z89	Q91z89 mus musculu
36	119	3.3	484	2	Q7R7B9	Q7r7b9 plasmodium
37	119	3.3	931	2	Q6QI35	Q6qi35 rattus norv
38	119	3.3	1102	2	Q6TUD6	Q6tud6 rattus norv
39	119	3.3	1300	2	P97692	P97692 rattus norv
40	119	3.3	1366	2	Q6QI58	Q6qi58 rattus norv
41	119	3.3	2321	2	Q6TXE2	Q6txe2 rattus norv
42	118	3.3	685	2	Q63288	Q63288 rattus norv
43	118	3.3	1217	2	Q6TUF6	Q6tuf6 rattus norv
44	118	3.3	1219	2	Q9QYM3	Q9qym3 mus musculu
45	118	3.3	1300	2	Q61785	Q61785 mus musculu

ALIGNMENTS

RESULT 1
Q9GZY1
ID Q9GZY1 PRELIMINARY; PRT; 135 AA.
AC Q9GZY1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RA Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1; -.
DR EMBL; AF189269; AAG17117.1; -.
DR Genew; HGNC:21079; PBOV1.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B1983705D0 CRC64;

Alignment Scores:
Pred. No.: 2.48e-64 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.34% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-83 (1-2088) x Q9GZY1 (1-135)

Qy 99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

```
Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
              |||||||||||||||||||||||||||||||||||||||||||
Db        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
```

RESULT 2
O00372
ID O00372 PRELIMINARY; PRT; 1275 AA.
AC O00372;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93570; AAC51273.1; -.
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT_1; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149110 MW; 4711B3BC22F7674E CRC64;

Alignment Scores:

Pred. No.:	1.09e-11	Length:	1275
Score:	200.00	Matches:	64
Percent Similarity:	42.02%	Conservative:	15
Best Local Similarity:	34.04%	Mismatches:	47
Query Match:	5.57%	Indels:	62
DB:	2	Gaps:	6

US-09-974-546C-83 (1-2088) x O00372 (1-1275)

```
Qy       1523 GTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAACCTTATGAGCAGGAAGAGGGAA 1582
              ||||||         |||   :::|||::::::||||||         ||||||:::|||
Db        294 ValCys---------ArgGlyLysPheThrAlaLeuAsnAlaTyrLysArgLysGlnGlu 310

Qy       1583 GGCTGTTTGAGAACCATTTACTTAGCAGAACCACATATTTTAGACACTTCCCTGCATTAA 1642
                                    ::::::         |||   :::   |||
Db        311 --------------------ArgSerLysIleAspThrLeuThrSerGlnLeuLysGlu 323

Qy       1643 CTGCACAAACAATATGTTTGCAAACTTGTTRGATCAACCTCCAACAACGACACATTCAGG 1702
              |||   ||||||                                          |||
Db        324 LeuGluLysGln--------------------------------GluGlnThrHis--- 331

Qy       1703 AGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCTAGAGATTGCTACAAATC 1762

Db        331 ------------------------------------------------------------ 331

Qy       1763 TTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTCCCTGTTATAAGCAGCAA 1822
                        |||   ||||||      :::||||||:::
Db        332 -----------SerLysAlaSerArgArgGlnGluIleThr----------------- 341

Qy       1823 GACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACATTCTTTCCAAAAGGAAC 1882
                  |||:::         |||                          |||   |||   |||
Db        342 ---LysIleArgAlaGluLeuLysGluIleGluThrGlnLysThrLeuGlnLysIleAsn 360

Qy       1883 TCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAACCATCTGAAGAAACTTCC 1942
                                       |||   ::::::         |||      :::
Db        361 --------------------------GluSerArgSerTrpPhePheGluArgIleAsn 371

Qy       1943 CAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGAC 2002
              ::::::                          ||||||||||||||||||||||||||
Db        372 LysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAsp 391

Qy       2003 ACAATAAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 2062
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 ThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrThrIle 411

Qy       2063 AGAGAATACTACAAACACCTCTAC 2086
              |||||||||||||||||||||||
Db        412 ArgGluTyrTyrLysHisLeuTyr 419
```

RESULT 3
O00375
ID O00375 PRELIMINARY; PRT; 1275 AA.
AC O00375;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93572; AAC51276.1; -.
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT_1; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149054 MW; 456DD0F3DD7A17F5 CRC64;

Alignment Scores:

Pred. No.:	1.57e-11	Length:	1275
Score:	198.50	Matches:	76
Percent Similarity:	37.74%	Conservative:	21
Best Local Similarity:	29.57%	Mismatches:	75

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:24:09 ; Search time 6016.29 Seconds
(without alignments)
16816.746 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 2088
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2086	99.9	2103	9	AF189270	AF189270 Homo sapi
	2	2076	99.4	2087	6	AR146835	AR146835 Sequence
	3	1961.6	93.9	2512	9	AF189269	AF189269 Homo sapi
	4	1950	93.4	2505	6	AR146836	AR146836 Sequence
	5	1920.2	92.0	129227	9	HS171N11	AL031433 Human DNA
	6	1920.2	92.0	186698	2	AC036236	AC036236 Homo sapi
	7	1919.2	91.9	2496	9	BC069109	BC069109 Homo sapi
	8	275.4	13.2	757	6	AR146757	AR146757 Sequence
	9	129.2	6.2	230685	2	AC131318	AC131318 Mus muscu
	10	123.2	5.9	167998	2	AC026382	AC026382 Mus muscu
	11	122	5.8	110000	2	AC114782_1	Continuation (2 of
	12	122	5.8	163271	9	AC093897	AC093897 Homo sapi
c	13	121	5.8	154941	9	AC117430	AC117430 Homo sapi
	14	121	5.8	155722	2	AC053487	AC053487 Homo sapi
	15	120.4	5.8	180388	9	HUMRETBLAS	L11910 Human retin
	16	120.4	5.8	203753	9	AC104343	AC104343 Homo sapi
c	17	119.8	5.7	108175	9	AC068206	AC068206 Homo sapi
	18	119.8	5.7	159299	9	AC090257	AC090257 Homo sapi
	19	119.8	5.7	169571	9	AC020687	AC020687 Homo sapi
	20	119.8	5.7	209152	2	AC015558	AC015558 Homo sapi
	21	119.6	5.7	167961	2	AP002422	AP002422 Homo sapi
c	22	119.4	5.7	95650	9	AC119737	AC119737 Homo sapi
	23	119.4	5.7	126838	9	AC125609	AC125609 Homo sapi
	24	119.4	5.7	169402	2	AC027715	AC027715 Homo sapi
	25	118.8	5.7	38528	9	HSAC001642	AC001642 Human Chr
	26	118.8	5.7	73589	9	AP001770	AP001770 Homo sapi
	27	118.8	5.7	109666	9	AP000838	AP000838 Homo sapi
c	28	118.8	5.7	158485	9	AC099545	AC099545 Homo sapi
	29	118.8	5.7	160426	2	AC068372	AC068372 Homo sapi
c	30	118.8	5.7	165870	9	AC025089	AC025089 Homo sapi
	31	118.8	5.7	174916	9	AC108056	AC108056 Homo sapi
c	32	118.8	5.7	175560	9	AL499602	AL499602 Human DNA
c	33	118.8	5.7	176802	2	AC012383	AC012383 Homo sapi
	34	118.8	5.7	184490	9	HSU82828	U82828 Homo sapien
	35	118.8	5.7	186104	9	AP005718	AP005718 Homo sapi
c	36	118.8	5.7	194149	9	AC100799	AC100799 Homo sapi
	37	118.8	5.7	196695	2	AC078936	AC078936 Homo sapi
	38	118.8	5.7	241802	2	AC147280	AC147280 Pan trogl
	39	117.8	5.6	161478	9	AC025538	AC025538 Homo sapi
	40	117.8	5.6	200923	9	AC091928	AC091928 Homo sapi
c	41	117.2	5.6	40392	9	AC004610	AC004610 Homo sapi
c	42	117.2	5.6	49743	9	AL450466	AL450466 Human DNA
c	43	117.2	5.6	51407	9	AL357556	AL357556 Human DNA
	44	117.2	5.6	74625	9	AL596214	AL596214 Human DNA
	45	117.2	5.6	78607	9	AC145092	AC145092 Pan trogl

ALIGNMENTS

RESULT 1
AF189270
LOCUS AF189270 2103 bp mRNA linear PRI 22-MAR-2001
DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION AF189270
VERSION AF189270.1 GI:10441603
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2103)
AUTHORS An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
Wright,G.L. Jr. and Veltri,R.W.
TITLE Cloning and characterization of UROC28, a novel gene overexpressed
in prostate, breast, and bladder cancers
JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PUBMED 11156405
REFERENCE 2 (bases 1 to 2103)
AUTHORS An,G. and Veltri,R.W.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
Parkway, Oklahoma City, OK 73104, USA
FEATURES Location/Qualifiers
source 1. .2103
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
gene 1. .2103
/gene="UC28"
CDS 99. .506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17118.1"
/db_xref="GI:10441604"
/translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS

```
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     misc_feature    1961. .2103
                     /gene="UC28"
                     /note="alternative splicing"
     polyA_signal    2047. .2052
                     /gene="UC28"
ORIGIN

  Query Match             99.9%;  Score 2086;  DB 9;  Length 2103;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2087;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980

Qy       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040

Qy       2041 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||| |||||||||||||||||||||||||||||||||||||
Db       2041 CACAGAAATAAAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088


RESULT 2
AR146835
LOCUS       AR146835               2087 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 83 from patent US 6218529.
ACCESSION   AR146835
VERSION     AR146835.1  GI:15110024
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2087)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 83 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .2087
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             99.4%;  Score 2076;  DB 6;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2087;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1979

Qy       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1980 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2039

Qy       2041 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2040 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2087


RESULT 3
AF189269
LOCUS       AF189269                2512 bp    mRNA    linear   PRI 22-MAR-2001
DEFINITION  Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION   AF189269
VERSION     AF189269.1  GI:10441601
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2512)
  AUTHORS   An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
            Wright,G.L. Jr. and Veltri,R.W.
  TITLE     Cloning and characterization of UROC28, a novel gene overexpressed
            in prostate, breast, and bladder cancers
  JOURNAL   Cancer Res. 60 (24), 7014-7020 (2000)
  MEDLINE   21028101
   PUBMED   11156405
REFERENCE   2  (bases 1 to 2512)
  AUTHORS   An,G. and Veltri,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
            Parkway, Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1. .2512
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q23-q24"
                     /tissue_type="prostate cancer"
     gene            1. .2512
                     /gene="UC28"
     CDS             99. .506
                     /gene="UC28"
                     /codon_start=1
                     /product="UC28 protein"
                     /protein_id="AAG17117.1"
                     /db_xref="GI:10441602"
                     /translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
                     LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     polyA_signal    2487. .2492
                     /gene="UC28"
ORIGIN
```

Query Match 93.9%; Score 1961.6; DB 9; Length 2512;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
```

```
Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||||||    |  |||  ||  | | |
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978


RESULT 4
AR146836
LOCUS       AR146836                2505 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 85 from patent US 6218529.
ACCESSION   AR146836
VERSION     AR146836.1  GI:15110025
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2505)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 85 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .2505
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             93.4%;  Score 1950;  DB 6;  Length 2505;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1967;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
```

```
Qy       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy      1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy      1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy      1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy      1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy      1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
```

```
             ||||||||||||||||||||||||||||||||||||||||   |  |||  ||  | |
Db      1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1977
```

```
RESULT 5
HS171N11
LOCUS       HS171N11              129227 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
            complete sequence.
ACCESSION   AL031433
VERSION     AL031433.4  GI:4826442
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129227)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 13, 1999 this sequence version replaced gi:4582115.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-171N11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .129227
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q23"
                     /clone="RP1-171N11"
                     /clone_lib="RPCI-1"
ORIGIN

  Query Match             92.0%;  Score 1920.2;  DB 9;  Length 129227;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 1961;  Conservative    1;  Mismatches   14;  Indels    3;  Gaps    3;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      29796 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 29855
Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
Db      29856 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 29913
Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
Db      29914 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 29973
Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
Db      29974 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 30033
Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
Db      30034 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 30093
Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
Db      30094 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 30153
Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
Db      30154 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 30213
Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
Db      30214 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 30273
Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
Db      30274 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 30333
Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
Db      30334 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 30393
Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
Db      30394 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 30453
Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
Db      30454 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 30513
Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
Db      30514 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 30573
Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
Db      30574 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 30633
Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db      30634 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 30693
Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db      30694 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 30753
Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db      30754 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 30813
Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db      30814 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 30873
Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db      30874 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 30933

Qy       1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
Db      30934 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 30993
Qy       1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
Db      30994 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 31053
Qy       1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
Db      31054 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 31113
Qy       1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
Db      31114 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 31173
Qy       1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
Db      31174 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 31233
Qy       1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
Db      31234 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 31293
Qy       1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
Db      31294 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 31353
Qy       1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
Db      31354 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 31413
Qy       1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
Db      31414 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 31473
Qy       1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
Db      31474 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 31533
Qy       1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
Db      31534 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 31593
Qy       1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
Db      31594 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 31653
Qy       1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
Db      31654 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 31713
Qy       1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
Db      31714 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 31772
```

```
RESULT 6
AC036236
LOCUS       AC036236              186698 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-133O15, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC036236
VERSION     AC036236.2  GI:9958166
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:7523932.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0133O15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183373 bases at least Q40
            Consensus quality: 184432 bases at least Q30
            Consensus quality: 185064 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 186098; sum-of-contigs
            Quality coverage: 5.84 in Q20 bases; agarose-fp
            Quality coverage: 5.57 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4769: contig of 4769 bp in length
            *     4770     4869: gap of unknown length
            *     4870    11657: contig of 6788 bp in length
            *    11658    11757: gap of unknown length
            *    11758    28189: contig of 16432 bp in length
            *    28190    28289: gap of unknown length
            *    28290    53341: contig of 25052 bp in length
            *    53342    53441: gap of unknown length
            *    53442   102327: contig of 48886 bp in length
            *   102328   102427: gap of unknown length
            *   102428   182781: contig of 80354 bp in length
            *   182782   182881: gap of unknown length
            *   182882   186698: contig of 3817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186698
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-133O15"
     misc_feature    1. .4769
                     /note="assembly_name:Contig10"
     misc_feature    4870. .11657
                     /note="assembly_name:Contig11"
     misc_feature    11758. .28189
                     /note="assembly_name:Contig12"
     misc_feature    28290. .53341
                     /note="assembly_name:Contig13"
     misc_feature    53442. .102327
                     /note="assembly_name:Contig14"
     misc_feature    102428. .182781
                     /note="assembly_name:Contig15
                     clone_end:SP6
                     vector_side:right"
     misc_feature    182882. .186698
                     /note="assembly_name:Contig9"
```

```
ORIGIN

  Query Match             92.0%;  Score 1920.2;  DB 2;  Length 186698;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 1961;  Conservative    1;  Mismatches   14;  Indels    3;  Gaps    3;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159156 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 159215

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              |||||||||||||||||||||||||||||| | | ||||||||||||||||||||||
Db     159216 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 159273

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159274 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 159333

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159334 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 159393

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159394 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 159453

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159454 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 159513

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159514 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 159573

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159574 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 159633

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159634 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 159693

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159694 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 159753

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159754 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 159813

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159814 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 159873

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159874 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 159933

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159934 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 159993

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159994 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 160053

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160054 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 160113

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160114 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 160173
```

```
Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160174 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 160233

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160234 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 160293

Qy       1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160294 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 160353

Qy       1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160354 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 160413

Qy       1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160414 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 160473

Qy       1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160474 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 160533

Qy       1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160534 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 160593

Qy       1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160594 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 160653

Qy       1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160654 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 160713

Qy       1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160714 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 160773

Qy       1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     160774 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 160833

Qy       1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160834 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 160893

Qy       1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160894 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 160953

Qy       1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160954 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 161013

Qy       1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161014 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 161073

Qy       1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             ||||||||||||||||||||||||||||||||||||||    |  |||  ||  | | |
Db     161074 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 161132
```

```
RESULT 7
BC069109
LOCUS       BC069109                2496 bp    mRNA    linear   PRI 23-JUN-2004
DEFINITION  Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cDNA
            clone IMAGE:7216926).
ACCESSION   BC069109
VERSION     BC069109.1  GI:46575743
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2496)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2496)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Anup Madan, University of Iowa
            cDNA Library Preparation: Anup Madan, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Neurogenomics Research Lab,
            200 B EMRB, University of Iowa, Iowa City, IA-52242
            anup-madan@uiowa.edu
            Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0.
FEATURES             Location/Qualifiers
     source          1. .2496
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7216926"
                     /tissue_type="Synthetic constructs"
                     /clone_lib="NIH_MGC_242"
                     /lab_host="XL10"
                     /note="Vector: pUC19"
ORIGIN
```

```
  Query Match             91.9%;  Score 1919.2;  DB 9;  Length 2496;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 1960;  Conservative    1;  Mismatches   14;  Indels    3;  Gaps    3;

Qy          2 ACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTAT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA- 59

Qy         62 TGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACCA 121
              |||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
Db         60 TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACCA 118
```

```
Qy       122 GAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAG 181
Db       119 GAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAG 178

Qy       182 ATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATT 241
Db       179 ATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATT 238

Qy       242 CTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTT 301
Db       239 CTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTT 298

Qy       302 CATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTT 361
Db       299 CATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTT 358

Qy       362 GACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCAC 421
Db       359 GACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCAC 418

Qy       422 ATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 481
Db       419 ATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 478

Qy       482 AACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAA 541
Db       479 AACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAA 538

Qy       542 TAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACT 601
Db       539 TAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACT 598

Qy       602 GAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTCA 661
Db       599 GAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTCA 658

Qy       662 ATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCTA 721
Db       659 ATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCTA 718

Qy       722 TTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTTT 781
Db       719 TTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTTT 778

Qy       782 CTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGATG 841
Db       779 CTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGATG 838

Qy       842 CTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTTA 901
Db       839 CTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTTA 898

Qy       902 GTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGTG 961
Db       899 GTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGTG 958

Qy       962 TAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATTT 1021
Db       959 TAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATTT 1018

Qy      1022 AATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTTG 1081
Db      1019 AATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTTG 1078

Qy      1082 TTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC- 1140
Db      1079 TTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATCA 1138

Qy      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
Db      1139 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1198

Qy      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
Db      1199 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1258

Qy      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
Db      1259 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1318

Qy      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
Db      1319 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1378

Qy      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
Db      1379 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1438

Qy      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
Db      1439 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1498

Qy      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
Db      1499 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1558

Qy      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
Db      1559 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1618

Qy      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
Db      1619 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAAC 1678

Qy      1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
Db      1679 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1738

Qy      1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
Db      1739 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1798

Qy      1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
Db      1799 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1858

Qy      1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
Db      1859 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1918

Qy      1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
Db      1919 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1976
```

```
RESULT 8
AR146757
LOCUS       AR146757                757 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 3 from patent US 6218529.
ACCESSION   AR146757
VERSION     AR146757.1  GI:15109946
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 3 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

```
  Query Match             13.2%;  Score 275.4;  DB 6;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 3e-42;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db             1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
              |||||||||||||||||||||||||||||||||   |  |||  ||  | |
Db           241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293


RESULT 9
AC131318
LOCUS       AC131318              230685 bp    DNA     linear   HTG 20-FEB-2004
DEFINITION  Mus musculus chromosome 10 clone RP23-367D2 map 10, *** SEQUENCING
            IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AC131318
VERSION     AC131318.4  GI:42716133
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 230685)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 10, clone RP23-367D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230685)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 230685)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 20, 2004 this sequence version replaced gi:29150479.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26968
                Center clone name: 367_D_2
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    47907: contig of 47907 bp in length
            *    47908    48007: gap of 100 bp
            *    48008    74241: contig of 26234 bp in length
            *    74242    74341: gap of 100 bp
            *    74342    89037: contig of 14696 bp in length
            *    89038    89137: gap of 100 bp
            *    89138   118227: contig of 29090 bp in length
            *   118228   118327: gap of 100 bp
            *   118328   158787: contig of 40460 bp in length
            *   158788   158887: gap of 100 bp
            *   158888   171882: contig of 12995 bp in length
            *   171883   171982: gap of 100 bp
            *   171983   196741: contig of 24759 bp in length
            *   196742   196841: gap of 100 bp
            *   196842   230685: contig of 33844 bp in length.
FEATURES             Location/Qualifiers
     source          1. .230685
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /map="10"
                     /clone="RP23-367D2"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

  Query Match             6.2%;  Score 129.2;  DB 2;  Length 230685;
  Best Local Similarity   59.3%;  Pred. No. 2.8e-15;
  Matches  416;  Conservative    1;  Mismatches  244;  Indels   41;  Gaps   10;

Qy       1194 GATCTATAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTA 1253
              ||  | |||||||| || ||||| || ||||        || |  || || |||||| ||
Db     125132 GAATTCTAGAACATTCACAGTAAATGCTGAG------AGCTTACATCTGTGCTTAACTTC 125185

Qy       1254 AGATTATGTGAGAAAAGTTTAAAG---ACACTTAGTAGAGTGATTTTGAAATATAGTAAA 1310
              |||  ||    | | ||||   ||   | | |    |      |  | | |||     |
Db     125186 TGATCATAGTGGGAGAGTTATTAGGTTAGAGTAGTTCTCAAGTATGTCCAATGGTACAGT 125245
```

```
Qy       1311 CACTTGGAAATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTC 1370
Db     125246 ACCTAGACATGGTGATACTTATGAAAAGAAGTTAATAGATAATATGGAAACCT-CATCTC 125304

Qy       1371 AAAAATAATGCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTC 1430
Db     125305 ACAAAGAGTGAATAAACGAAAGAAAAAATTTACAGTTATGGACTTCTGCATGTTTGCTTC 125364

Qy       1431 ATTACTTTTTCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAA 1490
Db     125365 TTCT------------TTATAATCATCTGCCTAAGGTAGTTATTTCAAAGGGTAAA-AGA 125411

Qy       1491 GATGGGAGGAAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCT 1550
Db     125412 GAAAGAAGGAAAATATAATTAGTATTGTATAAATGTGTGTGCCAGGTAGAAAAATCTTAT 125471

Qy       1551 CATCACTCAACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAG 1610
Db     125472 C-TCAGTCAACATTAGCAGCAGGAA---GGAAAGCTGCCTAAAAACAAATTACTTAGGAG 125527

Qy       1611 AACCACA---------TATTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTT 1661
Db     125528 AACCACACACGGTTGGTAGAGTATGCTGTAAGGTTGTTAGACTCAACCCCAACACCAGCA 125587

Qy       1662 GCAAACTTGTTRGAT-CAACCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATC 1720
Db     125588 GTTAAGGAGTTATATACAATTTCCAACAGCAGTACGCTGTAGACTCATACATTTTTCATC 125647

Qy       1721 AAACATTGGATTTTTCCTTAACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATG 1780
Db     125648 AAGCATTTGATTTCTCCTTAGCGCCAAAGACTATTATAAATCTTCCAAAGAATCGCTATG 125707

Qy       1781 GCTTCAGGCTAAGAAG----AGATTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCAT 1836
Db     125708 GCTTACGGTAAGAAGGGCTTGCTTTGCCCCTTGATATAAGGAGGTAAACAACTCAGTCAT 125767

Qy       1837 TTCACTCTCAAACTTCACTAATGATCACATTCTTTCCAAAAG 1878
Db     125768 TTTAGCCTCAAGCTTTGCTCATGATCATTCTTTCCATAGAAG 125809
```

```
RESULT 10
AC026382
LOCUS       AC026382               167998 bp    DNA     linear   HTG 16-OCT-2001
DEFINITION  Mus musculus clone RP23-246G10, WORKING DRAFT SEQUENCE, 19
            unordered pieces.
ACCESSION   AC026382
VERSION     AC026382.8  GI:16118081
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 167998)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167998)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:11094624.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MADA
                Center clone name: RP23-246G10
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 96% of reads
                Chemistry: Dye-terminator Big Dye: 4% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159501 bases at least Q40
                Consensus quality: 166426 bases at least Q30
                Consensus quality: 169264 bases at least Q20
                Estimated insert size: 169620; sum-of-contigs estimation
                Estimated insert size: 152000; agarose-fp estimation
                Quality coverage: 4.5x in Q20 bases; agarose-fp estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    29927: contig of 29927 bp in length
            *    29928    30027: gap of unknown length
            *    30028    56267: contig of 26240 bp in length
            *    56268    56367: gap of unknown length
            *    56368    69896: contig of 13529 bp in length
            *    69897    69996: gap of unknown length
            *    69997    83796: contig of 13800 bp in length
            *    83797    83896: gap of unknown length
            *    83897    93698: contig of 9802 bp in length
            *    93699    93798: gap of unknown length
            *    93799   102657: contig of 8859 bp in length
            *   102658   102757: gap of unknown length
            *   102758   113241: contig of 10484 bp in length
            *   113242   113341: gap of unknown length
            *   113342   120111: contig of 6770 bp in length
            *   120112   120211: gap of unknown length
            *   120212   128808: contig of 8597 bp in length
            *   128809   128908: gap of unknown length
            *   128909   134501: contig of 5593 bp in length
            *   134502   134601: gap of unknown length
            *   134602   140411: contig of 5810 bp in length
            *   140412   140511: gap of unknown length
            *   140512   144731: contig of 4220 bp in length
            *   144732   144831: gap of unknown length
            *   144832   149476: contig of 4645 bp in length
            *   149477   149576: gap of unknown length
            *   149577   153295: contig of 3719 bp in length
            *   153296   153395: gap of unknown length
            *   153396   156120: contig of 2725 bp in length
            *   156121   156220: gap of unknown length
            *   156221   159545: contig of 3325 bp in length
            *   159546   159645: gap of unknown length
            *   159646   162911: contig of 3266 bp in length
            *   162912   163011: gap of unknown length
            *   163012   165637: contig of 2626 bp in length
            *   165638   165737: gap of unknown length
            *   165738   167998: contig of 2261 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .167998
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-246G10"
ORIGIN

  Query Match             5.9%;  Score 123.2;  DB 2;  Length 167998;
  Best Local Similarity   58.7%;  Pred. No. 4.3e-14;
  Matches  412;  Conservative    1;  Mismatches  249;  Indels   40;  Gaps   10;

Qy       1194 GATCTATAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTA 1253
             ||  | ||||||||  || |||||  || ||||        || |  || || ||||||  ||
Db      51613 GAATTCTAGAACATTCACAGTAAATGCTGAG------AGCTTACATCTGTGCTTAACTTC 51666

Qy       1254 AGATTATGTGAGAAAAGTTTAAAG---ACACTTAGTAGAGTGATTTTGAAATATAGTAAA 1310
              ||| ||    | | ||||   ||   | | |   |     |  | |  |||      |
Db      51667 TGATCATAGTGGGAGAGTTATTAGGTTAGAGTAGTTCTCAAGTATGTCCAATGGTACAGT 51726

Qy       1311 CACTTGGAAATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTC 1370
               || |  | |        | | ||||||  ||||||||||||||| ||| || ||||||
Db      51727 ACCTAGACATGGTGATACTTATGAAAAGAAGTTAATAGATAATATGGAAACCT-CATCTC 51785

Qy       1371 AAAAATAATGCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTC 1430
              | ||| | || |||||| |   |||  |    ||  | |  |     | ||   ||  |||
Db      51786 ACAAAGAGTGAATAAACGAAAGAAAAAATTTACAGTTATGGACTTCTGCATGTTTGCTTC 51845

Qy       1431 ATTACTTTTTCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAA 1490
               |              || | | |||||||  |||| ||| || ||||||||||||   |
Db      51846 TTCT-----------TTATAATCATCTGCCTAAGGTAGTTATTTCAAAGGGTAAAATGA 51893

Qy       1491 GATGGG-AGGAAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTC 1549
              ||  |  |||||||    | | || ||| ||| | | ||| ||| | |||||||||  ||
Db      51894 GAAAGTAAGGAAAATATAATTAGTATTGTATAAATGTGTGTGCCAGGTAGAAAAATCTTA 51953

Qy       1550 TCATCACTCAACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCA 1609
              || ||| ||||| |||  |||||||||    |||| ||||   | | ||| | |||||||| |
Db      51954 TC-TCAGTCAACATTAGCAGCAGGAA---GGAAAGCTGCCTAAAAACAAATTACTTAGGA 52009

Qy       1610 GAAC-CACATATTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACT 1668
              |||| ||||   ||   | | |   ||| |        | |  |||        | | |
Db      52010 GAACACACACGGTTGGTAGAGTATGCTGTAAGGTTGTTAGACTCAACCCCAACACCAGCA 52069

Qy       1669 TGTTRG--------ATCAACCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATC 1720
               | :|          |||  ||||||| |   ||  |   || | | | ||||||||||
Db      52070 GTTAAGGAGTTATATACAATTTCCAACAGCAGTACGCTGTAGACTCATACATTTTTCATC 52129

Qy       1721 AAACATTGGATTTTTCCTTAACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATG 1780
              || |||| ||||| |||||| ||| | ||| |  || ||||||||  |||  || | |||
Db      52130 AAGCATTTGATTTCTCCTTAGCGCCAAAGACTATTATAAATCTTCCAAAGAATCGCTATG 52189

Qy       1781 GCTTCAGGCTAAGAAGAGA----TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCAT 1836
              ||||  ||  |  | |       || | || || |||||| ||  | |||| | || |||
Db      52190 GCTTACGGTAAGAAGGGCTTGCTTTGCCCCTTGATATAAGGAGGTAAACAACTCAGTCAT 52249

Qy       1837 TTCACTCTCAAACTTCACTAATGATCACATTCTTTCCAAAAG 1878
              || |  ||||| |||  || ||||||||   | |     | |||
Db      52250 TTTAGCCTCAAGCTTTGCTCATGATCATTCTTTCCATAGAAG 52291


RESULT 11
AC114782_1
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC114782 Accession AC114782
     Fragment Name      Begin     End
     AC114782_0             1    110000
     AC114782_1        100001    210000
     AC114782_2        200001    310000
     AC114782_3        300001    410000
     AC114782_4        400001    474678
Continuation (2 of 5) of AC114782 from base 100001 (AC114782 Homo sapiens chromosome UNK

  Query Match             5.8%;  Score 122;  DB 2;  Length 110000;
  Best Local Similarity   89.7%;  Pred. No. 8.1e-14;
  Matches  131;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1943 CAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGAC 2002
              |||   | | |  | ||  ||| ||  ||| | ||||||||||||||||||||||||||
Db      54002 CAAAATTGATAGACCGCTAGCAAGATGACAAAGAAAAAAAGAGAGAAGAATCAAATAGAC 54061

Qy       2003 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 2062
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54062 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 54121

Qy       2063 AGAGAATACTACAAACACCTCTACGC 2088
              |||||||||||||||||||||||||
Db      54122 AGAGAATACTACAAACACCTCTACGC 54147


RESULT 12
AC093897
LOCUS       AC093897              163271 bp    DNA     linear   PRI 01-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-721G13 from 4, complete sequence.
ACCESSION   AC093897 AC067901
VERSION     AC093897.3  GI:16304304
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163271)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 163271)
  AUTHORS   Van Brunt,A. and Kozlowicz,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-721G13
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 163271)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 163271)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 163271)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 21, 2001 this sequence version replaced gi:15825639.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0721G13
                Drafting Center: WIBR
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is AC019264. Actual start of this clone is at base position 1 of RP11-721G13; actual end is at base position 163271 of RP11-721G13.

Data from AC019264 and AC093652 was used to finish this clone AC093897. Polymorphisms have been identified between AC093652 and AC093897. Single plasmid region exists between 149738 and 149749.

The sequence of AC067901 has been incorporated into AC093897.

FEATURES Location/Qualifiers
source 1. .163271
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-721G13"
/clone_lib="RPCI-11"
repeat_region 4. .213
/rpt_family="L2"
repeat_region 759. .841
/rpt_family="MIR"
repeat_region 1069. .1278
/rpt_family="MIR"
repeat_region 1362. .1502
/rpt_family="MIR"
repeat_region 1822. .1879
/rpt_family="(TA)n"
repeat_region 2408. .2568
/rpt_family="MIR"
repeat_region 2793. .2971
/rpt_family="MIR"
repeat_region 3003. .3177
/rpt_family="MIR"
repeat_region 3182. .3395
/rpt_family="MER1_type"
repeat_region 3733. .3757
/rpt_family="AT_rich"
repeat_region 4460. .4510
/rpt_family="CR1"
repeat_region 5168. .5230
/rpt_family="(TA)n"
repeat_region 5475. .5496
/rpt_family="AT_rich"
repeat_region 6246. .6333
/rpt_family="MaLR"
repeat_region 6886. .7074
/rpt_family="MaLR"
repeat_region 7177. .7326
/rpt_family="MIR"
repeat_region 8674. .8984
/rpt_family="Alu"
repeat_region 8792. .8824
/rpt_family="(CA)n"
repeat_region 9027. .9055
/rpt_family="AT_rich"
repeat_region 9048. .9329
/rpt_family="L1"
repeat_region 9473. .9742
/rpt_family="L1"
repeat_region 9757. .10020
/rpt_family="L1"
repeat_region 10230. .10278
/rpt_family="AT_rich"
repeat_region 10470. .10492
/rpt_family="AT_rich"
repeat_region 10471. .10748
/rpt_family="L1"
repeat_region 11055. .11253
/rpt_family="MaLR"
misc_feature 12547. .12839
/note="similar to EST AA677976 (NID:g2658498) zi14c09.s1"
repeat_region 12845. .13026
/rpt_family="MER1_type"
repeat_region 12957. .12989
/rpt_family="AT_rich"
repeat_region 13588. .14502
/rpt_family="MER1_type"
repeat_region 13962. .13982
/rpt_family="AT_rich"
repeat_region 14495. .14643
/rpt_family="MER1_type"
repeat_region 14808. .15206
/rpt_family="MaLR"
repeat_region 15394. .15686
/rpt_family="L1"
repeat_region 15729. .15969
/rpt_family="MER2_type"
repeat_region 16051. .16096
/rpt_family="MER2_type"
repeat_region 16098. .16512
/rpt_family="L1"
repeat_region 16512. .16541
/rpt_family="AT_rich"
repeat_region 16799. .16960
/rpt_family="L1"
misc_feature 17129. .17628
/note="similar to EST BF815983 (NID:g12151333)"
repeat_region 21028. .21056
/rpt_family="AT_rich"
repeat_region 21356. .21578
/rpt_family="L1"
repeat_region 22771. .23095
/rpt_family="L2"
repeat_region 23221. .23249
/rpt_family="(TG)n"
repeat_region 23269. .23418
/rpt_family="Alu"
repeat_region 23432. .23455
/rpt_family="AT_rich"
repeat_region 23471. .23600
/rpt_family="Alu"
repeat_region 23583. .23616
/rpt_family="(A)n"
repeat_region 23732. .23864
/rpt_family="MIR"
repeat_region 23894. .24261
/rpt_family="MaLR"
repeat_region 25230. .25252
/rpt_family="AT_rich"
misc_feature 25492. .25655
/note="similar to EST BE833497 (NID:g10265875)"
repeat_region 25709. .25732

```
  Query Match             5.8%;  Score 122;  DB 9;  Length 163271;
  Best Local Similarity   89.7%;  Pred. No. 7.3e-14;
  Matches  131;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1943 CAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGAC 2002
             |||   | | |  | ||  ||| ||  ||| | ||||||||||||||||||||||||||
Db      40437 CAAAATTGATAGACCGCTAGCAAGATGACAAAGAAAAAAAGAGAGAAGAATCAAATAGAC 40496

Qy       2003 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 2062
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40497 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 40556

Qy       2063 AGAGAATACTACAAACACCTCTACGC 2088
              |||||||||||||||||||||||||
Db      40557 AGAGAATACTACAAACACCTCTACGC 40582


RESULT 13
AC117430/c
LOCUS       AC117430               154941 bp    DNA     linear   PRI 27-JUN-2002
DEFINITION  Homo sapiens 3 BAC RP11-119D18 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC117430
VERSION     AC117430.3  GI:20340433
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154941)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154941)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 154941)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 154941)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 154941)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 30, 2002 this sequence version replaced gi:20279244.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES              Location/Qualifiers
     source           1. .154941
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="3"
                      /clone="RP11-119D18"
     misc_feature     1. .2000
                      /note="overlaps bases 113280. .115279 of clone AC112506"
```

```
                /function="clone overlap"
repeat_region   1. .538
                /rpt_family="L1M3b"
repeat_region   910. .1310
                /rpt_family="L2"
repeat_region   2848. .2899
                /rpt_family="(TA)n"
repeat_region   3958. .3979
                /rpt_family="AT_rich"
repeat_region   complement(5020. .5227)
                /rpt_family="MIR"
repeat_region   6721. .6843
                /rpt_family="LTR48"
repeat_region   complement(12729. .13025)
                /rpt_family="AluY"
repeat_region   13784. .13843
                /rpt_family="(TA)n"
repeat_region   complement(14404. .14713)
                /rpt_family="AluSg"
repeat_region   14716. .14744
                /rpt_family="AT_rich"
repeat_region   14807. .14861
                /rpt_family="MIR"
repeat_region   14862. .15148
                /rpt_family="AluSc"
repeat_region   15149. .15255
                /rpt_family="MIR"
repeat_region   15680. .16066
                /rpt_family="L1MA2"
repeat_region   16915. .17029
                /rpt_family="FLAM_C"
repeat_region   17052. .17097
                /rpt_family="FLAM_A"
repeat_region   17098. .17452
                /rpt_family="THE1B"
repeat_region   17453. .17538
                /rpt_family="FLAM_A"
repeat_region   17760. .17790
                /rpt_family="AT_rich"
repeat_region   17943. .17975
                /rpt_family="T-rich"
repeat_region   complement(17977. .18298)
                /rpt_family="AluJo"
repeat_region   18947. .18967
                /rpt_family="(TTTA)n"
repeat_region   complement(18970. .24273)
                /rpt_family="L1PB1"
repeat_region   complement(24298. .25342)
                /rpt_family="L1PB1"
repeat_region   25434. .25475
                /rpt_family="T-rich"
repeat_region   26099. .26122
                /rpt_family="AT_rich"
repeat_region   complement(26363. .26501)
                /rpt_family="L1MA2"
repeat_region   26674. .26985
                /rpt_family="AluSg"
repeat_region   27131. .27200
                /rpt_family="AT_rich"
repeat_region   complement(27385. .28146)
                /rpt_family="L1MA4"
repeat_region   28219. .28301
                /rpt_family="A-rich"
repeat_region   complement(28751. .29047)
                /rpt_family="AluJb"
repeat_region   complement(29048. .29125)
                /rpt_family="MIR"
repeat_region   29505. .29812
                /rpt_family="LTR16A"
repeat_region   complement(30217. .30329)
                /rpt_family="MER8"
repeat_region   complement(30330. .30426)
                /rpt_family="HERV16"
```

```
     repeat_region   30572. .30742

  Query Match             5.8%;  Score 121;  DB 9;  Length 154941;
  Best Local Similarity   78.4%;  Pred. No. 1.1e-13;
  Matches  145;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1904 AGTTAAGAACATCAAAACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGC 1963
             | ||||  |   ||  |   |   | | | | |   |   |||   | | |  ||||  ||
Db     147147 AATTAATGAATCCAGGAGCTAGTTTTTTGAAAGGATCAACAAAATTGATAAACTGCTAGC 147088

Qy       1964 ACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGG 2023
             | ||| |   |  ||||||||||||||||||||||||||||||||||||||||||||||||
Db     147087 AAGACTAATAAAGAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGG 147028

Qy       2024 GATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTC 2083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147027 GATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTC 146968

Qy       2084 TACGC 2088
             |||||
Db     146967 TACGC 146963
```

```
RESULT 14
AC053487
LOCUS       AC053487               155722 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-119D18, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC053487
VERSION     AC053487.3  GI:9958145
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155722)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155722)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:9795672.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0119D18
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 76%
            Sequencing vector: plasmid; 24%
            Chemistry: Dye-primer ET; 74% of reads
            Chemistry: Dye-terminator Big Dye; 26% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 154612 bases at least Q40
            Consensus quality: 154798 bases at least Q30
            Consensus quality: 154903 bases at least Q20
            Insert size: 132000; agarose-fp
            Insert size: 154934; sum-of-contigs
            Quality coverage: 8.60 in Q20 bases; agarose-fp
            Quality coverage: 7.32 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   154934: contig of 154934 bp in length
            *    154935   155034: gap of unknown length
            *    155035   155722: contig of 688 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155722
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-119D18"
     misc_feature    1. .154934
                     /note="assembly_name:Contig43
                     clone_end:SP6
                     vector_side:right"
     misc_feature    155035. .155722
                     /note="assembly_name:Contig15"
ORIGIN

  Query Match             5.8%;  Score 121;  DB 2;  Length 155722;
  Best Local Similarity   78.4%;  Pred. No. 1.1e-13;
  Matches  145;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1904 AGTTAAGAACATCAAAACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGC 1963
             | ||||  |   ||  |   |   | |  | |   |   |||   | | |  |||| ||
Db       7795 AATTAATGAATCCAGGAGCTAGTTTTTTGAAAGGATCAACAAAATTGATAAACTGCTAGC 7854

Qy       1964 ACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGG 2023
             | ||| |   |  ||||||||||||||||||||||||||||||||||||||||||||||
Db       7855 AAGACTAATAAAGAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGG 7914

Qy       2024 GATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTC 2083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7915 GATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTC 7974

Qy       2084 TACGC 2088
             |||||
Db       7975 TACGC 7979


RESULT 15
HUMRETBLAS
LOCUS       HUMRETBLAS            180388 bp    DNA     linear   PRI 23-NOV-1994
DEFINITION  Human retinoblastoma susceptibility gene exons 1-27, complete cds.
ACCESSION   L11910
VERSION     L11910.1  GI:292420
KEYWORDS    nuclear protein; recessive oncogene; retinoblastoma gene;
            retinoblastoma protein; retinoblastoma susceptibility; tumor
            supressor gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Friend,S.H., Bernards,R., Rogelj,S., Weinberg,R.A., Rapaport,J.M.,
            Albert,D.M. and Dryja,T.P.
  TITLE     A human DNA segment with properties of the gene that predisposes to
            retinoblastoma and osteosarcoma
  JOURNAL   Nature 323 (6089), 643-646 (1986)
  MEDLINE   87039336
   PUBMED   2877398
REFERENCE   2  (sites)
  AUTHORS   Friend,S.H., Horowitz,J.M., Gerber,M.R., Wang,X.F., Bogenmann,E.,
            Li,F.P. and Weinberg,R.A.
  TITLE     Deletions of a DNA sequence in retinoblastomas and mesenchymal
            tumors: organization of the sequence and its encoded protein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (24), 9059-9063 (1987)
  MEDLINE   88097427
   PUBMED   3480530
REFERENCE   3  (sites)
  AUTHORS   Lee,W.H., Bookstein,R., Hong,F., Young,L.J., Shew,J.Y. and Lee,E.Y.
  TITLE     Human retinoblastoma susceptibility gene: cloning, identification,
            and sequence
  JOURNAL   Science 235 (4794), 1394-1399 (1987)
  MEDLINE   87149066
   PUBMED   3823889
REFERENCE   4  (sites)
  AUTHORS   McGee,T.L., Yandell,D.W. and Dryja,T.P.
  TITLE     Structure and partial genomic sequence of the human retinoblastoma
            susceptibility gene
  JOURNAL   Gene 80 (1), 119-128 (1989)
  MEDLINE   90006771
   PUBMED   2701949
REFERENCE   5  (sites)
  AUTHORS   Sakai,T., Ohtani,N., McGee,T.L., Robbins,P.D. and Dryja,T.P.
  TITLE     Oncogenic germ-line mutations in Sp1 and ATF sites in the human
            retinoblastoma gene
  JOURNAL   Nature 353 (6339), 83-86 (1991)
  MEDLINE   91351319
   PUBMED   1881452
REFERENCE   6  (bases 1 to 180388)
  AUTHORS   Toguchida,J., McGee,T.L., Paterson,J.C., Eagle,J.R., Tucker,S.,
            Yandell,D.W. and Dryja,T.P.
  TITLE     Complete genomic sequence of the human retinoblastoma
            susceptibility gene
  JOURNAL   Genomics 17 (3), 535-543 (1993)
  MEDLINE   94063891
   PUBMED   7902321
COMMENT     Original source text: Homo sapiens DNA.
FEATURES             Location/Qualifiers
     source          1. .180388
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     protein_bind    1858. .1863
                     /bound_moiety="SP1"
                     /evidence=experimental
     protein_bind    1866. .1872
                     /bound_moiety="ATF"
                     /evidence=experimental
     CDS             join(2060. .2196,5424. .5550,39446. .39561,41926. .42045,
                     44668. .44706,45799. .45866,56853. .56963,59651. .59793,
                     61730. .61807,64330. .64439,65364. .65441,70242. .70329,
                     73753. .73869,76430. .76486,76889. .76920,77001. .77077,
                     78083. .78279,149998. .150116,153208. .153353,156693. .156838,
                     160730. .160834,161997. .162110,162204. .162367,
                     170372. .170402,173707. .173849,174361. .174410,
                     177005. .177078)
                     /codon_start=1
                     /product="retinoblastoma susceptibility protein"
                     /protein_id="AAA53483.1"
                     /db_xref="GI:292421"
                     /translation="MPPKTPRKTAATAAAAAAEPPAPPPPPPPEEDPEQDSGPEDLPL
                     VRLEFEETEEPDFTALCQKLKIPDHVRERAWLTWEKVSSVDGVLGGYIQKKKELWGIC
                     IFIAAVDLDEMSFTFTELQKNIEISVHKFFNLLKEIDTSTKVDNAMSRLLKKYDVLFA
                     LFSKLERTCELIYLTQPSSSISTEINSALVLKVSWITFLLAKGEVLQMEDDLVISFQL
                     MLCVLDYFIKLSPPMLLKEPYKTAVIPINGSPRTPRRGQNRSARIAKQLENDTRIIEV
                     LCKEHECNIDEVKNVYFKNFIPFMNSLGLVTSNGLPEVENLSKRYEEIYLKNKDLDAR
                     LFLDHDKTLQTDSIDSFETQRTPRKSNLDEEVNVIPPHTPVRTVMNTIQQLMMILNSA
                     SDQPSENLISYFNNCTVNPKESILKRVKDIGYIFKEKFAKAVGQGCVEIGSQRYKLGV
                     RLYYRVMESMLKSEEERLSIQNFSKLLNDNIFHMSLLACALEVVMATYSRSTSQNLDS
                     GTDLSFPWILNVLNLKAFDFYKVIESFIKAEGNLTREMIKHLERCEHRIMESLAWLSD
                     SPLFDLIKQSKDREGPTDHLESACPLNLPLQNNHTAADMYLSPVRSPKKKGSTTRVNS
                     TANAETQATSAFQTQKPLKSTSLSLFYKKVYRLAYLRLNTLCERLLSEHPELEHIIWT
                     LFQHTLQNEYELMRDRHLDQIMMCSMYGICKVKNIDLKFKIIVTAYKDLPHAVQETFK
                     RVLIKEEEYDSIIVFYNSVFMQRLKTNILQYASTRPPTLSPIPHIPRSPYKFPSSPLR
                     IPGGNIYISPLKSPYKISEGLPTPTKMTPRSRILVSIGESFGTSEKFQKINQMVCNSD
                     RVLKRSAEGSNPPKPLKKLRFDIEGSDEADGSKHLPGESKFQQKLAEMTSTRTRMQKQ
                     KMNDSMDTSNKEEK"
     exon            <2060. .2196
                     /number=1
     intron          2197. .5423
                     /number=1
     variation       2302
                     /standard_name="Restriction Fragment Length Polymorphism"
```

```
                /note="BamHI RFLP"
     exon       5424. .5550
                /number=2
     intron     5551. .39445
                /number=2
     satellite  14868. .14914
                /standard_name="microsatellite RBi2 (AFM058xd6)"
     exon       39446. .39561
                /number=3
     intron     39562. .41925
                /number=3
     exon       41926. .42045
                /number=4
     intron     42046. .44667
                /number=4
     satellite  43218. .43257
                /standard_name="microsatellite RBi4"
     exon       44668. .44706
                /number=5
     intron     44707. .45798
                /number=5
     exon       45799. .45866
                /number=6
     intron     45867. .56852
                /number=6
     exon       56853. .56963
                /number=7
     intron     56964. .59650
                /number=7
     exon       59651. .59793
                /number=8
     intron     59794. .61729
                /number=8
     exon       61730. .61807
                /number=9
     intron     61808. .64329
                /number=9
     exon       64330. .64439
                /number=10
     intron     64440. .65363
                /number=10
     exon       65364. .65441
                /number=11
     intron     65442. .70241
                /number=11
     exon       70242. .70329
                /number=12
     intron     70330. .73752
                /number=12
     exon       73753. .73869
                /number=13
     intron     73870. .76429
                /number=13
     exon       76430. .76486
                /number=14
     intron     76487. .76888
                /number=14
     exon       76889. .76920
                /number=15
     intron     76921. .77000
                /number=15
     exon       77001. .77077
                /number=16
     intron     77078. .78082
                /number=16
     exon       78083. .78279
                /number=17
     intron     78280. .149997
                /number=17
     variation  99426
                /standard_name="Restriction Fragment Length Polymorphism"
                /note="XbaI RFLP"
     satellite  102921. .102959
                /standard_name="microsatellite RBi17"
     repeat_region  123912. .125501
                /standard_name="Variable Number Tandem Repeats (VNTR)"
                /rpt_type=tandem
     exon       149998. .150116
                /number=18
     intron     150117. .153207
                /number=18
     exon       153208. .153353
                /number=19
     intron     153354. .156692
                /number=19
     exon       156693. .156838
                /number=20
     intron     156839. .160729
                /number=20
     satellite  156895. .156976
                /standard_name="microsatellite RB1.20"
     exon       160730. .160834
                /number=21
     intron     160835. .161996
                /number=21
     exon       161997. .162110
                /number=22
```

```
  Query Match             5.8%;  Score 120.4;  DB 9;  Length 180388;
  Best Local Similarity   89.0%;  Pred. No. 1.4e-13;
  Matches  130;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1943 CAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGAC 2002
              |||   | | |  | ||  ||| ||| |   | ||||||||||||||||||||||||
Db     165647 CAAAATTGATAGACCGCTAGCAAGACTAATAAGAAAAAAAGAGAGAAGAATCAAATAGAC 165706

Qy       2003 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 2062
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165707 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 165766

Qy       2063 AGAGAATACTACAAACACCTCTACGC 2088
              |||||||||||||||||||||||||
Db     165767 AGAGAATACTACAAACACCTCTACGC 165792


Search completed: September  9, 2005, 03:30:32
Job time : 6025.29 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 23:15:40 ; Search time 241.343 Seconds
(without alignments)
14156.413 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 2088
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2076	99.4	2087	3	US-09-097-199-83	Sequence 83, Appl
	2	1950	93.4	2505	3	US-09-097-199-85	Sequence 85, Appl
	3	1920.2	92.0	2506	4	US-09-949-016-3311	Sequence 3311, Ap
	4	1920.2	92.0	6507	4	US-09-949-016-15053	Sequence 15053, A
c	5	572.2	27.4	601	4	US-09-949-016-119365	Sequence 119365,
c	6	572.2	27.4	601	4	US-09-949-016-119366	Sequence 119366,
c	7	344.2	16.5	601	4	US-09-949-016-119364	Sequence 119364,
	8	275.4	13.2	757	2	US-08-692-787-3	Sequence 3, Appli
	9	275.4	13.2	757	3	US-09-097-199-3	Sequence 3, Appli
	10	118.8	5.7	146401	4	US-09-949-016-16151	Sequence 16151, A
c	11	117.2	5.6	162841	4	US-09-949-016-13733	Sequence 13733, A
c	12	115.6	5.5	573	3	US-09-385-982-420	Sequence 420, App
c	13	115.2	5.5	430	3	US-09-397-787-248	Sequence 248, App
c	14	115.2	5.5	475	4	US-09-621-976-12318	Sequence 12318, A
c	15	115.2	5.5	32042	3	US-09-245-281-44	Sequence 44, Appl
c	16	115.2	5.5	32042	4	US-09-340-620A-63	Sequence 63, Appl
	17	115.2	5.5	35837	4	US-09-949-016-15232	Sequence 15232, A
	18	115.2	5.5	86439	4	US-09-949-016-11945	Sequence 11945, A
	19	115.2	5.5	86440	4	US-09-949-016-16990	Sequence 16990, A
	20	115.2	5.5	99370	4	US-09-949-016-12816	Sequence 12816, A
	21	115.2	5.5	99370	4	US-09-949-016-17540	Sequence 17540, A
c	22	115.2	5.5	102520	4	US-09-949-016-17367	Sequence 17367, A
c	23	115.2	5.5	102526	4	US-09-949-016-12448	Sequence 12448, A
c	24	115.2	5.5	116592	4	US-09-818-512-3	Sequence 3, Appli
c	25	115.2	5.5	124264	4	US-09-949-016-16396	Sequence 16396, A
	26	115.2	5.5	168394	4	US-09-949-016-13002	Sequence 13002, A
c	27	115.2	5.5	175236	4	US-09-949-016-14353	Sequence 14353, A
c	28	115.2	5.5	206433	4	US-09-949-016-13527	Sequence 13527, A
c	29	115.2	5.5	254778	4	US-09-949-016-12417	Sequence 12417, A
c	30	115.2	5.5	312470	4	US-09-949-016-14043	Sequence 14043, A
c	31	115.2	5.5	336024	4	US-09-949-016-12373	Sequence 12373, A
	32	115.2	5.5	363032	4	US-09-949-016-12415	Sequence 12415, A
	33	115.2	5.5	363033	4	US-09-949-016-15754	Sequence 15754, A
c	34	113.6	5.4	439	3	US-09-397-787-269	Sequence 269, App
	35	113.6	5.4	601	4	US-09-949-016-116730	Sequence 116730,
	36	113.6	5.4	4656	4	US-09-949-016-14956	Sequence 14956, A
c	37	113.6	5.4	69813	4	US-09-949-016-12455	Sequence 12455, A
c	38	113.6	5.4	69813	4	US-09-949-016-13905	Sequence 13905, A
c	39	113.6	5.4	69813	4	US-09-949-016-13906	Sequence 13906, A
c	40	113.6	5.4	69833	4	US-09-949-016-12861	Sequence 12861, A
c	41	113.6	5.4	80706	4	US-09-949-016-15347	Sequence 15347, A
c	42	113.6	5.4	87648	4	US-09-949-016-13655	Sequence 13655, A
c	43	113.6	5.4	95648	4	US-09-949-016-13139	Sequence 13139, A
c	44	113.6	5.4	106380	4	US-09-949-016-17553	Sequence 17553, A
	45	113.6	5.4	107421	4	US-09-949-016-15532	Sequence 15532, A

ALIGNMENTS

```
RESULT 1
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  83:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2087 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..503
US-09-097-199-83

  Query Match             99.4%;  Score 2076;  DB 3;  Length 2087;
```

```
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2087;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAA 1979

Qy       1981 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1980 AAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATATCACCACCGATCC 2039

Qy       2041 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       2040 CACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2087
```

RESULT 2
US-09-097-199-85

; Sequence 85, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..503
US-09-097-199-85

Query Match 93.4%; Score 1950; DB 3; Length 2505;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
```

```
Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1977
```

```
RESULT 3
US-09-949-016-3311
; Sequence 3311, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3311
;   LENGTH: 2506
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-3311

  Query Match             92.0%;  Score 1920.2;  DB 4;  Length 2506;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 1961;  Conservative    1;  Mismatches   14;  Indels    3;  Gaps    3;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              |||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
```

```
Db         61 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 118

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 178

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 238

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 298

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 358

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 418

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 478

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 538

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 598

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 658

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 718

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 778

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 838

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 898

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 958

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1018

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1078

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1138

Qy       1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1198
```

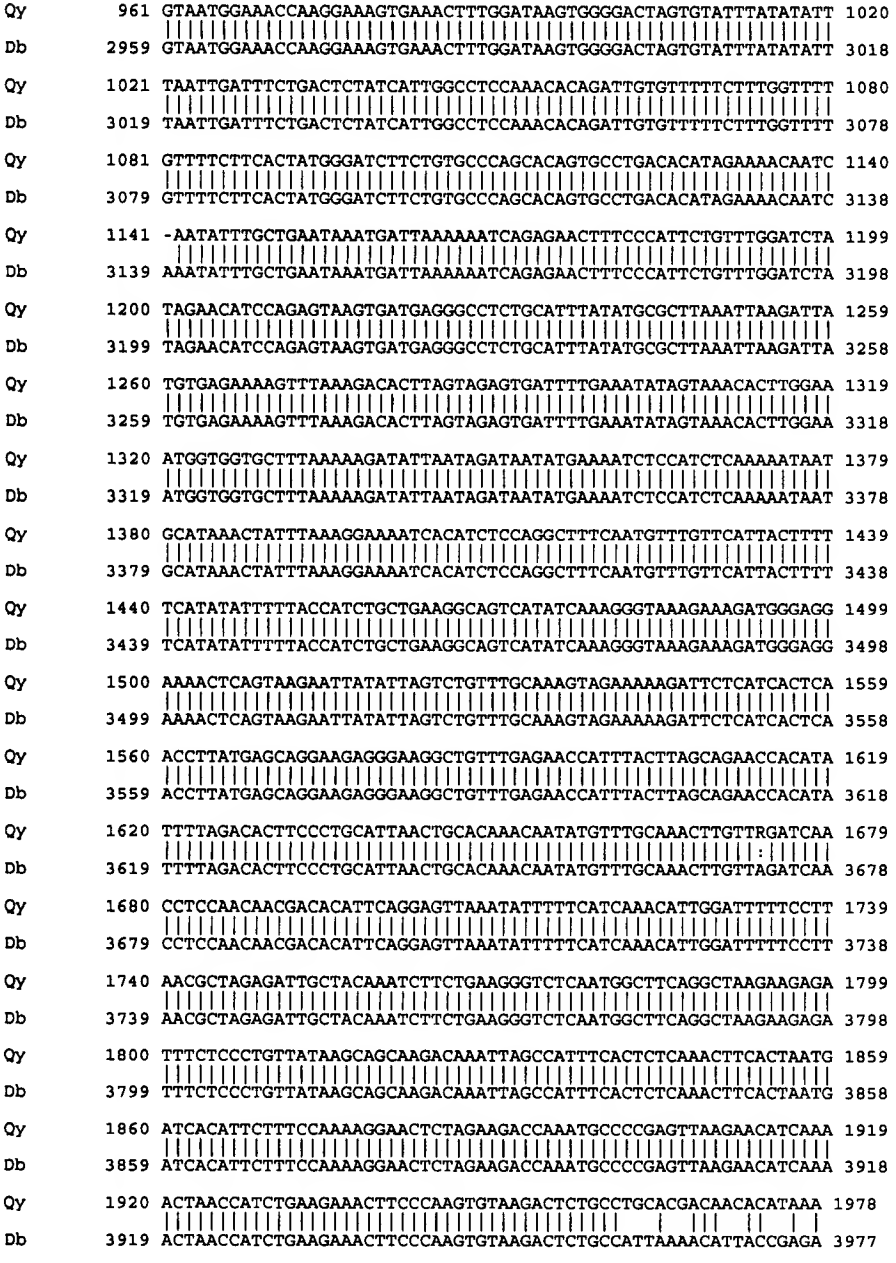
```
Qy       1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1258

Qy       1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1318

Qy       1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1378

Qy       1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1438

Qy       1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1498

Qy       1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1558

Qy       1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1618

Qy       1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db       1619 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 1678

Qy       1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1738

Qy       1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1798

Qy       1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1858

Qy       1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1918

Qy       1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||||||   |  |||  ||  | | |
Db       1919 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1977
```

RESULT 4
US-09-949-016-15053
; Sequence 15053, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 15053
; LENGTH: 6507
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15053

Query Match 92.0%; Score 1920.2; DB 4; Length 6507;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1961; Conservative 1; Mismatches 14; Indels 3; Gaps 3;

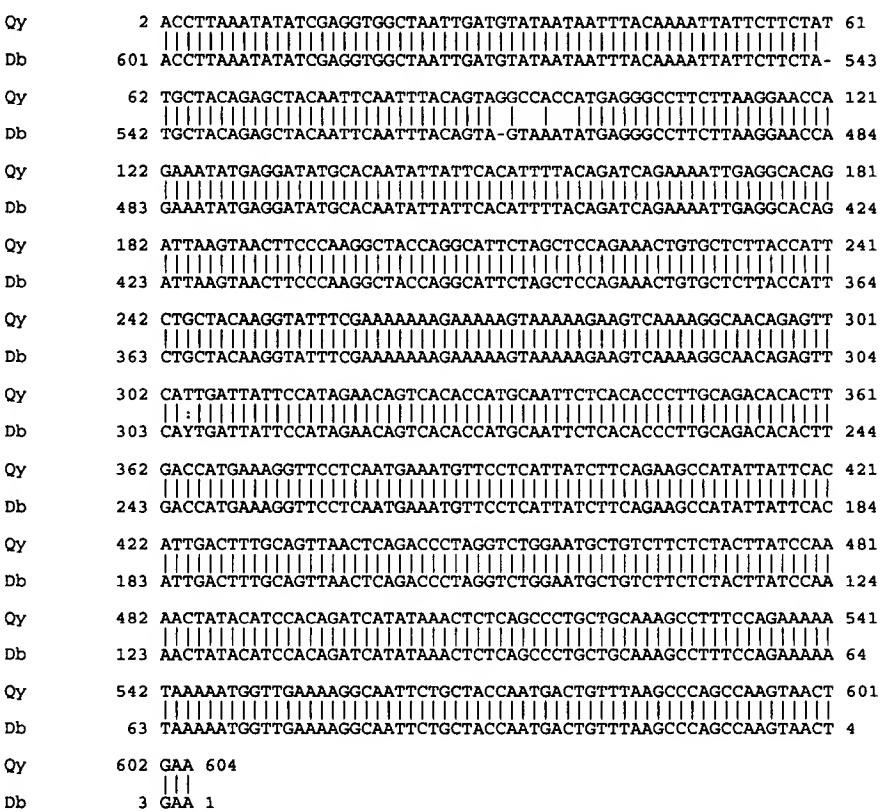
```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2001 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 2060

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              |||||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
Db       2061 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 2118

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2119 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 2178

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2179 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 2238

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2239 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 2298

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2299 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 2358

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2359 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 2418

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2419 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 2478

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2479 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 2538

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2539 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 2598

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2599 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 2658

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2659 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 2718

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2719 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 2778

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2779 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 2838

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2839 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 2898

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2899 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 2958
```



RESULT 5
US-09-949-016-119365/c
; Sequence 119365, Application US/09949016

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119365
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119365

Query Match 27.4%; Score 572.2; DB 4; Length 601;
Best Local Similarity 98.8%; Pred. No. 1.6e-132;
Matches 596; Conservative 1; Mismatches 4; Indels 2; Gaps 2;



RESULT 6
US-09-949-016-119366/c

```
; Sequence 119366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119366
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119366

  Query Match             27.4%;  Score 572.2;  DB 4;  Length 601;
  Best Local Similarity   98.8%;  Pred. No. 1.6e-132;
  Matches  596;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    2;

Qy          14 TCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTATTGCTACAGAGCT 73
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         601 TCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA-TGCTACAGAGCT 543

Qy          74 ACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGA 133
               |||||||||||||||||  |   |  ||||||||||||||||||||||||||||||||||
Db         542 ACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGA 484

Qy         134 TATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAGATTAAGTAACTT 193
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         483 TATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAGATTAAGTAACTT 424

Qy         194 CCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATTCTGCTACAAGGT 253
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         423 CCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATTCTGCTACAAGGT 364

Qy         254 ATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTTCATTGATTATTC 313
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         363 ATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTTCATTGATTATTC 304

Qy         314 CATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTTGACCATGAAAGG 373
               ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         303 CAYAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTTGACCATGAAAGG 244

Qy         374 TTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCACATTGACTTTGCA 433
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         243 TTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCACATTGACTTTGCA 184

Qy         434 GTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAAAACTATACATCC 493
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         183 GTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAAAACTATACATCC 124

Qy         494 ACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAATAAAAATGGTTG 553
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         123 ACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAATAAAAATGGTTG 64

Qy         554 AAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACTGAACCATTCCAA 613
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          63 AAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACTGAACCATTCCAA 4

Qy         614 CTT 616
               |||
Db           3 CTT 1


RESULT 7
US-09-949-016-119364/c
; Sequence 119364, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119364
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119364

  Query Match             16.5%;  Score 344.2;  DB 4;  Length 601;
  Best Local Similarity   98.1%;  Pred. No. 7.7e-76;
  Matches  368;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    2;

Qy           1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         373 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 314

Qy          61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
               |||||||||||||:|||||||||||||||||  |   |  ||||||||||||||||||||
Db         313 -TGCTACAGAGCTRCAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 256

Qy         121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         255 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 196

Qy         181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         195 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 136

Qy         241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         135 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 76

Qy         301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          75 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 16

Qy         361 TGACCATGAAAGGTT 375
               |||||||||||||||
Db          15 TGACCATGAAAGGTT 1


RESULT 8
US-08-692-787-3
; Sequence 3, Application US/08692787
; Patent No. 5882864
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE
;    TITLE OF INVENTION:  DISEASE
;    NUMBER OF SEQUENCES:  82
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
```

```
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  United States of America
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/692,787
;       FILING DATE:  Concurrently Herewith
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Corder, Timothy S.
;       REGISTRATION NUMBER:  38,414
;       REFERENCE/DOCKET NUMBER:  UROC:012
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  757 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-692-787-3

  Query Match             13.2%;  Score 275.4;  DB 2;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 1.1e-58;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293


RESULT 9
US-09-097-199-3
; Sequence 3, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  757 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-097-199-3

  Query Match             13.2%;  Score 275.4;  DB 3;  Length 757;
  Best Local Similarity   96.2%;  Pred. No. 1.1e-58;
  Matches  282;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
             |||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 293


RESULT 10
US-09-949-016-16151
; Sequence 16151, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16151
;   LENGTH: 146401
;   TYPE: DNA
```

```
;   ORGANISM: Human
US-09-949-016-16151

  Query Match             5.7%;  Score 118.8;  DB 4;  Length 146401;
  Best Local Similarity   88.4%;  Pred. No. 7.3e-19;
  Matches  129;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       1943 CAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGAC 2002
             |||   | | |  | ||  ||| ||| |   |  ||||||||||||||||||||||||||
Db     138898 CAAAATTGATAGACCGCTAGCAAGACTAATAAAGAAAAAAGAGAGAAGAATCAAATAGAC 138957

Qy       2003 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 2062
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138958 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 139017

Qy       2063 AGAGAATACTACAAACACCTCTACGC 2088
             |||||||||||||||||||||||||||
Db     139018 AGAGAATACTACAAACACCTCTACGC 139043


RESULT 11
US-09-949-016-13733/c
; Sequence 13733, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13733
;   LENGTH: 162841
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13733

  Query Match             5.6%;  Score 117.2;  DB 4;  Length 162841;
  Best Local Similarity   87.7%;  Pred. No. 1.9e-18;
  Matches  128;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       1943 CAAGTGTAAGACTCTGCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGAC 2002
             |||   | | |  ||||  ||| ||| |   |  ||||||||||||||||||||||||||
Db     155543 CAAAATTGATAGACTGCTAGCAAGACTAATAAAGAAAAAAGAGAGAAGAATCAAATAGAC 155484

Qy       2003 ACAATAAAAAATGATAAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATC 2062
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     155483 GCAATAAAAAATGATAAAGGGGATATCACCACCAATCCCACAGAAATACAAACTACCATC 155424

Qy       2063 AGAGAATACTACAAACACCTCTACGC 2088
             |||||||||||||||||||||||||||
Db     155423 AGAGAATACTACAAACACCTCTACGC 155398


RESULT 12
US-09-385-982-420/c
; Sequence 420, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 420
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(573)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-420

  Query Match             5.5%;  Score 115.6;  DB 3;  Length 573;
  Best Local Similarity   95.2%;  Pred. No. 5.1e-19;
  Matches  118;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1965 CGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGG 2024
             | | |  | |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        312 CAAGACTANATAAAAAAAGAGAGAAGAATCAAATAGACNCAATAAAAAATGATAAAGGGG 253

Qy       2025 ATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCT 2084
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCT 193

Qy       2085 ACGC 2088
             ||||
Db        192 ACGC 189


RESULT 13
US-09-397-787-248/c
; Sequence 248, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C2
;  CURRENT APPLICATION NUMBER: US/09/397,787
;  CURRENT FILING DATE: 1999-09-16
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 248
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(430)
;   OTHER INFORMATION: n = A,T,C or G
US-09-397-787-248

  Query Match             5.5%;  Score 115.2;  DB 3;  Length 430;
  Best Local Similarity   97.5%;  Pred. No. 5.7e-19;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
             || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 AATAAAGAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 236

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 176
```

```
RESULT 14
US-09-621-976-12318/c
; Sequence 12318, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 12318
;   LENGTH: 475
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-12318

  Query Match             5.5%;  Score 115.2;  DB 4;  Length 475;
  Best Local Similarity   97.5%;  Pred. No. 5.9e-19;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 AATAAAGAAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 85

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 25


RESULT 15
US-09-245-281-44/c
; Sequence 44, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 07334/118001
;  CURRENT APPLICATION NUMBER: US/09/245,281
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: US 09/207,359
;  EARLIER FILING DATE: 1998-12-08
;  EARLIER APPLICATION NUMBER: US 09/099,041
;  EARLIER FILING DATE: 1998-06-17
;  EARLIER APPLICATION NUMBER: US 09/019,942
;  EARLIER FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
;   LENGTH: 32042
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-245-281-44

  Query Match             5.5%;  Score 115.2;  DB 3;  Length 32042;
  Best Local Similarity   97.5%;  Pred. No. 3.1e-18;
  Matches  117;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1969 AACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 2028
              || | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15038 AATAAAGAAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGGGATAT 14979

Qy       2029 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 2088
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14978 CACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTCTACGC 14919


Search completed: September  9, 2005, 06:27:38
```

Job time : 401.343 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 9, 2005, 03:30:51 ; Search time 1108.99 Seconds
(without alignments)
12364.714 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 2088
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2076	99.4	2087	10	US-09-974-546-83	Sequence 83, Appl
	2	1950	93.4	2505	10	US-09-974-546-85	Sequence 85, Appl
	3	275.4	13.2	757	10	US-09-974-546-3	Sequence 3, Appli
c	4	115.6	5.5	573	10	US-09-871-161-420	Sequence 420, App
c	5	115.2	5.5	430	9	US-09-876-889-248	Sequence 248, App
	6	115.2	5.5	4180	11	US-09-984-429-334	Sequence 334, App
	7	115.2	5.5	4906	16	US-10-082-828A-143	Sequence 143, App
	8	115.2	5.5	4906	24	US-11-057-807-143	Sequence 143, App
	9	115.2	5.5	6007	21	US-10-278-698-245	Sequence 245, App
	10	115.2	5.5	6007	21	US-10-278-698-759	Sequence 759, App
	11	115.2	5.5	6009	11	US-09-984-429-339	Sequence 339, App
	12	115.2	5.5	6052	11	US-09-984-429-425	Sequence 425, App
	13	115.2	5.5	6052	11	US-09-984-429-447	Sequence 447, App
	14	115.2	5.5	6075	21	US-10-800-322-32	Sequence 32, Appl
	15	115.2	5.5	6539	14	US-10-025-201-1	Sequence 1, Appli
	16	115.2	5.5	7626	13	US-10-001-835-82	Sequence 82, Appl
	17	115.2	5.5	11661	11	US-09-984-429-345	Sequence 345, App
c	18	115.2	5.5	30515	9	US-09-764-847-1208	Sequence 1208, Ap
c	19	115.2	5.5	30515	14	US-10-092-154-1208	Sequence 1208, Ap
c	20	115.2	5.5	32042	9	US-09-728-721-63	Sequence 63, Appl
c	21	115.2	5.5	32042	13	US-10-118-984-44	Sequence 44, Appl
c	22	115.2	5.5	32042	15	US-10-295-981-63	Sequence 63, Appl
c	23	115.2	5.5	32042	21	US-10-843-188-63	Sequence 63, Appl
c	24	115.2	5.5	34001	15	US-10-006-883A-15	Sequence 15, Appl
	25	115.2	5.5	68571	17	US-10-401-194-1	Sequence 1, Appli
c	26	115.2	5.5	86500	20	US-10-719-993-6771	Sequence 6771, Ap
c	27	115.2	5.5	116592	9	US-09-818-512-3	Sequence 3, Appli
c	28	115.2	5.5	116592	15	US-10-354-065-3	Sequence 3, Appli
c	29	115.2	5.5	116592	24	US-11-074-646-3	Sequence 3, Appli
c	30	115.2	5.5	147300	22	US-10-723-681-3	Sequence 3, Appli
	31	115.2	5.5	227931	17	US-10-085-117-274	Sequence 274, App
c	32	115.2	5.5	227931	17	US-10-085-117-274	Sequence 274, App
	33	115.2	5.5	383432	22	US-10-737-082-34	Sequence 34, Appl
	34	115.2	5.5	383432	22	US-10-765-790-34	Sequence 34, Appl
c	35	115.2	5.5	402850	10	US-09-844-653-5	Sequence 5, Appli
c	36	115.2	5.5	561515	19	US-10-741-601-5682	Sequence 5682, Ap
c	37	115.2	5.5	561515	21	US-10-741-600-17730	Sequence 17730, A
	38	115.2	5.5	606398	20	US-10-719-993-6782	Sequence 6782, Ap
	39	115.2	5.5	1503841	9	US-09-795-668-1	Sequence 1, Appli
c	40	115.2	5.5	1503841	9	US-09-795-668-1	Sequence 1, Appli
	41	115.2	5.5	1503841	9	US-09-795-686-1	Sequence 1, Appli
c	42	115.2	5.5	1503841	9	US-09-795-686-1	Sequence 1, Appli
	43	115.2	5.5	1503841	9	US-09-946-807-1	Sequence 1, Appli
c	44	115.2	5.5	1503841	9	US-09-946-807-1	Sequence 1, Appli
c	45	114.4	5.5	611	13	US-10-027-632-204073	Sequence 204073,

ALIGNMENTS

RESULT 1
US-09-974-546-83
; Sequence 83, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199

```
;             FILING DATE: 1998-06-12
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;   INFORMATION FOR SEQ ID NO: 83:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2087 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  99..503
;        SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-974-546-83

  Query Match             99.4%;  Score 2076;  DB 10;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2087;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:58:29 ; Search time 4510.06 Seconds
(without alignments)
17622.418 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 2088
Sequence: 1 gaccttaaatatatcgaggt.....tactacaaacacctctacgc 2088

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	119.8	5.7	412	8	AQ045191	AQ045191 RPCI11-34
c	2	119	5.7	460	8	AQ829037	AQ829037 HS_5023_B
	3	116.8	5.6	456	5	BQ353836	BQ353836 PM1-HT034
c	4	116.4	5.6	275	2	BE065285	BE065285 RC1-BT031
	5	115.2	5.5	335	5	BQ378810	BQ378810 IL2-UT007
	6	115.2	5.5	353	2	BF833793	BF833793 RC1-HT088
	7	115.2	5.5	383	1	AA776006	AA776006 ae82h10.s
	8	115.2	5.5	396	2	BE155551	BE155551 PM2-HT035
c	9	115.2	5.5	410	2	BE155550	BE155550 PM2-HT035
c	10	115.2	5.5	411	2	BE155549	BE155549 PM2-HT035
c	11	115.2	5.5	416	8	AQ337692	AQ337692 HS_5017_A
	12	115.2	5.5	466	5	BQ303951	BQ303951 RC1-BT031
c	13	115.2	5.5	474	8	AZ694823	AZ694823 UP_516-21
c	14	115.2	5.5	476	2	BE065326	BE065326 RC1-BT031
	15	115.2	5.5	478	2	BE154146	BE154146 PM1-HT034
	16	115.2	5.5	480	2	BE065400	BE065400 RC1-BT031
	17	115.2	5.5	493	5	BQ303952	BQ303952 RC1-BT031
c	18	115.2	5.5	498	8	B76374	B76374 RPCI11-15D1
c	19	115.2	5.5	559	8	B54683	B54683 CIT-HSP-384
	20	115.2	5.5	563	2	AW605861	AW605861 MR0-HT024
	21	115.2	5.5	573	5	BQ317875	BQ317875 RC0-CT047
c	22	115.2	5.5	638	1	AL698542	AL698542 DKFZp686D
	23	115.2	5.5	645	2	BE065286	BE065286 RC1-BT031
	24	115.2	5.5	668	2	BE065409	BE065409 RC1-BT031
	25	115.2	5.5	670	1	AI133053	AI133053 HA1642 Hu
	26	115.2	5.5	687	2	BE065399	BE065399 RC1-BT031
c	27	115.2	5.5	703	8	B66360	B66360 CIT-HSP-200
	28	115.2	5.5	771	2	BE065219	BE065219 RC1-BT031
	29	115.2	5.5	790	1	AU137403	AU137403 AU137403
c	30	115.2	5.5	790	8	BZ600251	BZ600251 WHABA17TF
c	31	115.2	5.5	805	8	BZ604730	BZ604730 WHADF83TR
c	32	114.6	5.5	544	8	AQ600990	AQ600990 HS_5374_A
c	33	114.6	5.5	724	8	AQ355228	AQ355228 CITBI-E1-
c	34	114.6	5.5	796	8	BZ611286	BZ611286 WHACL88TR
	35	114	5.5	755	9	AG031791	AG031791 Pan trogl
	36	113.8	5.5	213	2	AW818026	AW818026 CM3-ST027
	37	113.8	5.5	582	5	BX504526	BX504526 DKFZp686M
c	38	113.8	5.5	800	8	BZ600022	BZ600022 WHABK34TF
	39	113.6	5.4	270	5	BQ358592	BQ358592 PM1-HT034
	40	113.6	5.4	365	2	AW938273	AW938273 QV0-DT004
	41	113.6	5.4	374	2	AW938274	AW938274 QV0-DT004
	42	113.6	5.4	391	2	BE154240	BE154240 PM1-HT034
c	43	113.6	5.4	404	8	AQ488905	AQ488905 RPCI-11-2
	44	113.6	5.4	415	5	BQ353785	BQ353785 PM1-HT034
c	45	113.6	5.4	423	1	AV732317	AV732317 AV732317

ALIGNMENTS

RESULT 1
AQ045191/c
LOCUS AQ045191 412 bp DNA linear GSS 14-APR-1999
DEFINITION RPCI11-34D16.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-34D16, genomic survey sequence.
ACCESSION AQ045191
VERSION AQ045191.1 GI:3314118
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 412)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .412
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7512759"
/db_xref="taxon:9606"
/clone="RPCI-11-34D16"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
ORIGIN

Query Match 5.7%; Score 119.8; DB 8; Length 412;

```
  Best Local Similarity   94.7%;  Pred. No. 6.7e-16;
  Matches  124;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1958 GCCTGCACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGAT 2017
              | ||||  ||  ||  ||||||||||||||||||||||||||||||||||||||||||||
Db        306 GACTGCTAGAAGACTAATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGAT 247

Qy       2018 AAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAA 2077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 AAAGGGGATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAA 187

Qy       2078 CACCTCTACGC 2088
              ||||||||||||
Db        186 CACCTCTACGC 176


RESULT 2
AQ829037/c
LOCUS       AQ829037                460 bp    DNA     linear   GSS 27-AUG-1999
DEFINITION  HS_5023_B2_F06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=599 Col=12 Row=L, genomic survey sequence.
ACCESSION   AQ829037
VERSION     AQ829037.1  GI:5795099
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 599  row: L  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=599 Col=12 Row=L"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             5.7%;  Score 119;  DB 8;  Length 460;
  Best Local Similarity   77.3%;  Pred. No. 1e-15;
  Matches  143;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1904 AGTTAAGAACATCAAAACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGC 1963
              | | | |||       | ||   | | ||||   |     |  |   | | |  ||||  ||
Db        360 ATTAATGAATCCAGGAGCTGGTCTTTTGAAAGGATCAACATAAATTGATAGACTGCTAGC 301

Qy       1964 ACGACAACACATAAAAAAAGAGAGAAGAATCAAATAGACACAATAAAAAATGATAAAGGG 2023
              | ||| |   |  |||| ||||||||||||||||||||||||||| ||||||||||||
Db        300 AAGACTAATAAAGAAAAANAGAGAAGAATCAAATAGACACAATAANAAATGATAAAGGG 241

Qy       2024 GATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTC 2083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GATATCACCACCGATCCCACAGAAATACAAACTACCATCAGAGAATACTACAAACACCTC 181

Qy       2084 TACGC 2088
              |||||
Db        180 TACGC 176


RESULT 3
BQ353836
LOCUS       BQ353836                456 bp    mRNA    linear   EST 20-MAY-2002
DEFINITION  PM1-HT0340-150900-013-c12 HT0340 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ353836
VERSION     BQ353836.1  GI:21017892
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM1&t2=PM1-HT0340-
            150900-013-c12&t3=2000-09-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0340"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             5.6%;  Score 116.8;  DB 5;  Length 456;
  Best Local Similarity   98.3%;  Pred. No. 3.2e-15;
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:07:44 ; Search time 2414 Seconds
(without alignments)
2709.799 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174502_20945/app_query.fasta_1.654
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_2647_@runat_07092005_174502_20945 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	2087	6	AR146835	AR146835 Sequence
2	694	100.0	2103	9	AF189270	AF189270 Homo sapi
3	694	100.0	2496	9	BC069109	BC069109 Homo sapi
4	694	100.0	2505	6	AR146836	AR146836 Sequence
5	694	100.0	2512	9	AF189269	AF189269 Homo sapi
6	694	100.0	129227	9	HS171N11	AL031433 Human DNA
7	694	100.0	186698	2	AC036236	AC036236 Homo sapi
c 8	99.5	14.3	259920	2	AC096345	AC096345 Rattus no
9	95	13.7	9833	9	AF291597	AF291597 Homo sapi
10	94	13.5	9432	9	AF291608	AF291608 Homo sapi
11	94	13.5	9817	9	AF291631	AF291631 Homo sapi
12	94	13.5	9817	9	AF291632	AF291632 Homo sapi
13	94	13.5	9830	9	AF291644	AF291644 Homo sapi
14	94	13.5	9833	9	AF291587	AF291587 Homo sapi
15	94	13.5	9833	9	AF291588	AF291588 Homo sapi
16	94	13.5	9833	9	AF291589	AF291589 Homo sapi
17	94	13.5	9833	9	AF291590	AF291590 Homo sapi
18	94	13.5	9833	9	AF291591	AF291591 Homo sapi
19	94	13.5	9833	9	AF291592	AF291592 Homo sapi
20	94	13.5	9833	9	AF291593	AF291593 Homo sapi
21	94	13.5	9833	9	AF291594	AF291594 Homo sapi
22	94	13.5	9833	9	AF291595	AF291595 Homo sapi
23	94	13.5	9833	9	AF291598	AF291598 Homo sapi
24	94	13.5	9833	9	AF291599	AF291599 Homo sapi
25	94	13.5	9833	9	AF291600	AF291600 Homo sapi
26	94	13.5	9833	9	AF291601	AF291601 Homo sapi
27	94	13.5	9833	9	AF291602	AF291602 Homo sapi
28	94	13.5	9833	9	AF291603	AF291603 Homo sapi
29	94	13.5	9833	9	AF291604	AF291604 Homo sapi
30	94	13.5	9833	9	AF291605	AF291605 Homo sapi
31	94	13.5	9833	9	AF291606	AF291606 Homo sapi
32	94	13.5	9833	9	AF291607	AF291607 Homo sapi
33	94	13.5	9833	9	AF291609	AF291609 Homo sapi
34	94	13.5	9833	9	AF291610	AF291610 Homo sapi
35	94	13.5	9833	9	AF291612	AF291612 Homo sapi
36	94	13.5	9833	9	AF291614	AF291614 Homo sapi
37	94	13.5	9833	9	AF291615	AF291615 Homo sapi
38	94	13.5	9833	9	AF291618	AF291618 Homo sapi
39	94	13.5	9833	9	AF291620	AF291620 Homo sapi
40	94	13.5	9833	9	AF291621	AF291621 Homo sapi
41	94	13.5	9833	9	AF291623	AF291623 Homo sapi
42	94	13.5	9833	9	AF291624	AF291624 Homo sapi
43	94	13.5	9833	9	AF291625	AF291625 Homo sapi
44	94	13.5	9833	9	AF291626	AF291626 Homo sapi
45	94	13.5	9833	9	AF291627	AF291627 Homo sapi

ALIGNMENTS

RESULT 1
AR146835
LOCUS AR146835 2087 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 83 from patent US 6218529.
ACCESSION AR146835
VERSION AR146835.1 GI:15110024
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2087)
AUTHORS An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
TITLE Biomarkers and targets for diagnosis, prognosis and management of prostate, breast and bladder cancer
JOURNAL Patent: US 6218529-A 83 17-APR-2001;
FEATURES Location/Qualifiers
source 1. .2087
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Alignment Scores:

Pred. No.:	2.18e-67	Length:	2087
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

```
US-09-974-546C-84 (1-135) x AR146835 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
AF189270
LOCUS       AF189270                2103 bp    mRNA    linear   PRI 22-MAR-2001
DEFINITION  Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION   AF189270
VERSION     AF189270.1  GI:10441603
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2103)
  AUTHORS   An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
            Wright,G.L. Jr. and Veltri,R.W.
  TITLE     Cloning and characterization of UROC28, a novel gene overexpressed
            in prostate, breast, and bladder cancers
  JOURNAL   Cancer Res. 60 (24), 7014-7020 (2000)
  MEDLINE   21028101
   PUBMED   11156405
REFERENCE   2  (bases 1 to 2103)
  AUTHORS   An,G. and Veltri,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
            Parkway, Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1. .2103
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q23-q24"
                     /tissue_type="prostate cancer"
     gene            1. .2103
                     /gene="UC28"
     CDS             99. .506
                     /gene="UC28"
                     /codon_start=1
                     /product="UC28 protein"
                     /protein_id="AAG17118.1"
                     /db_xref="GI:10441604"
                     /translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
                     LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     misc_feature    1961. .2103
                     /gene="UC28"
                     /note="alternative splicing"
     polyA_signal    2047. .2052
                     /gene="UC28"
ORIGIN

Alignment Scores:
Pred. No.:                2.2e-67        Length:          2103
Score:                    694.00         Matches:         135
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-09-974-546C-84 (1-135) x AF189270 (1-2103)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 3
BC069109
LOCUS       BC069109                2496 bp    mRNA    linear   PRI 23-JUN-2004
DEFINITION  Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cDNA
            clone IMAGE:7216926).
ACCESSION   BC069109
VERSION     BC069109.1  GI:46575743
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2496)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
```

```
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2496)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Anup Madan, University of Iowa
            cDNA Library Preparation: Anup Madan, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Neurogenomics Research Lab,
            200 B EMRB, University of Iowa, Iowa City, IA-52242
            anup-madan@uiowa.edu
            Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0.
FEATURES                Location/Qualifiers
     source             1. .2496
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:7216926"
                        /tissue_type="Synthetic constructs"
                        /clone_lib="NIH_MGC_242"
                        /lab_host="XL10"
                        /note="Vector: pUC19"
ORIGIN

Alignment Scores:
Pred. No.:              2.66e-67        Length:         2496
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-974-546C-84 (1-135) x BC069109 (1-2496)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 155

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 215

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 275

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 335

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 395

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 455

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        456 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 500

RESULT 4
AR146836
LOCUS       AR146836               2505 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 85 from patent US 6218529.
ACCESSION   AR146836
VERSION     AR146836.1  GI:15110025
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2505)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 85 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .2505
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              2.68e-67        Length:         2505
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-974-546C-84 (1-135) x AR146836 (1-2505)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 5
AF189269
LOCUS       AF189269               2512 bp    mRNA    linear   PRI 22-MAR-2001
DEFINITION  Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION   AF189269
VERSION     AF189269.1  GI:10441601
KEYWORDS    .
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2512)
  AUTHORS   An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
            Wright,G.L. Jr. and Veltri,R.W.
  TITLE     Cloning and characterization of UROC28, a novel gene overexpressed
            in prostate, breast, and bladder cancers
  JOURNAL   Cancer Res. 60 (24), 7014-7020 (2000)
  MEDLINE   21028101
   PUBMED   11156405
REFERENCE   2  (bases 1 to 2512)
  AUTHORS   An,G. and Veltri,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
            Parkway, Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1. .2512
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q23-q24"
                     /tissue_type="prostate cancer"
     gene            1. .2512
                     /gene="UC28"
     CDS             99. .506
                     /gene="UC28"
                     /codon_start=1
                     /product="UC28 protein"
                     /protein_id="AAG17117.1"
                     /db_xref="GI:10441602"
                     /translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
                     LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     polyA_signal    2487. .2492
                     /gene="UC28"
ORIGIN

Alignment Scores:
Pred. No.:              2.68e-67        Length:         2512
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-974-546C-84 (1-135) x AF189269 (1-2512)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 6
HS171N11
LOCUS       HS171N11              129227 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
            complete sequence.
ACCESSION   AL031433
VERSION     AL031433.4  GI:4826442
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129227)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 13, 1999 this sequence version replaced gi:4582115.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-171N11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .129227
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q23"
                     /clone="RP1-171N11"
                     /clone_lib="RPCI-1"
ORIGIN

Alignment Scores:
Pred. No.:              2.19e-65        Length:         129227
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
```

```
DB:                    9              Gaps:          0

US-09-974-546C-84 (1-135) x HS171N11 (1-129227)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29892 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 29951

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29952 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 30011

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30012 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 30071

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30072 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 30131

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30132 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 30191

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30192 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 30251

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db      30252 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 30296

RESULT 7
AC036236
LOCUS       AC036236               186698 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-133O15, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC036236
VERSION     AC036236.2  GI:9958166
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:7523932.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0133O15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183373 bases at least Q40
            Consensus quality: 184432 bases at least Q30
            Consensus quality: 185064 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 186098; sum-of-contigs
            Quality coverage: 5.84 in Q20 bases; agarose-fp
            Quality coverage: 5.57 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4769: contig of 4769 bp in length
            *     4770     4869: gap of unknown length
            *     4870    11657: contig of 6788 bp in length
            *    11658    11757: gap of unknown length
            *    11758    28189: contig of 16432 bp in length
            *    28190    28289: gap of unknown length
            *    28290    53341: contig of 25052 bp in length
            *    53342    53441: gap of unknown length
            *    53442   102327: contig of 48886 bp in length
            *   102328   102427: gap of unknown length
            *   102428   182781: contig of 80354 bp in length
            *   182782   182881: gap of unknown length
            *   182882   186698: contig of 3817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186698
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-133O15"
     misc_feature    1. .4769
                     /note="assembly_name:Contig10"
     misc_feature    4870. .11657
                     /note="assembly_name:Contig11"
     misc_feature    11758. .28189
                     /note="assembly_name:Contig12"
     misc_feature    28290. .53341
                     /note="assembly_name:Contig13"
     misc_feature    53442. .102327
                     /note="assembly_name:Contig14"
     misc_feature    102428. .182781
                     /note="assembly_name:Contig15
                     clone_end:SP6
                     vector_side:right"
     misc_feature    182882. .186698
                     /note="assembly_name:Contig9"
ORIGIN

Alignment Scores:
Pred. No.:              3.3e-65        Length:          186698
Score:                  694.00         Matches:         135
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     2              Gaps:            0

US-09-974-546C-84 (1-135) x AC036236 (1-186698)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159252 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 159311

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159312 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 159371

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159372 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 159431

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     159432 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 159491

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159492 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 159551

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159552 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 159611

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||
Db     159612 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 159656
```

RESULT 8
AC096345/c
LOCUS AC096345 259920 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-18J15, WORKING DRAFT SEQUENCE, 2 unordered pieces.
ACCESSION AC096345
VERSION AC096345.6 GI:30521870
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 259920)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 259920)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 259920)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 10, 2003 this sequence version replaced gi:24818617.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GEYU
    Center clone name: CH230-18J15
-------------- Summary Statistics
    Assembly program: Atlas 3.0;
    Consensus quality: 240867 bases at least Q40
    Consensus quality: 244203 bases at least Q30
    Consensus quality: 245971 bases at least Q20
    Estimated insert size: 252807; sum-of-contigs estimation
    Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1   258395: contig of 258395 bp in length
*   258396   258495: gap of unknown length
*   258496   259920: contig of 1425 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .259920
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-18J15"
     misc_feature    1. .1739
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    8054. .8862
                     /note="clone_boundary
                     clone_end:T7
```

```
                     site:EcoRI
                     end_sequence:BH359975"
     misc_feature    complement(248170. .249015)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH360032"
     misc_feature    249451. .251486
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    256461. .258395
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              98.4          Length:         259920
Score:                  99.50         Matches:        38
Percent Similarity:     40.54%        Conservative:   22
Best Local Similarity:  25.68%        Mismatches:     41
Query Match:            14.34%        Indels:         47
DB:                     2             Gaps:           8

US-09-974-546C-84 (1-135) x AC096345 (1-259920)

Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              :::||||||:::         |||:::      |||:::|||      ||||||   |||
Db     195470 TTTGAGGATGTGCAGAGGACGATATACTTTTCACAACTAAGGTTAGAAAGACATCATCTG 195411

Qy         30 SerAsnPheProArgLeuProGlyIle------------LeuAlaProGluThrValLeu 45
              |||      |||       |||   :::                   |||:::   ::::::
Db     195410 TCCACAATGCCACATTTTCCAGAAGTAGGAAGACTGACATTTCAGCCAAAGAGAATTATT 195351

Qy         46 ---------------LeuProPheCysTyr---------------------------- 50
                           |||||||||   |||
Db     195350 TCCAAATTCAGTTTCCTCCCTTTCCTGTATCCAGCTCACTAATTCCCACATATTTCTTCT 195291

Qy         51 ------------------LysValPheArgLysLys------------GluLysValLys 60
                              :::|||   |||:::                  |||:::
Db     195290 TTCCTGCCTTCTGTTTCAAATCTCTTTTGGAAAAGAACTGAGTAGGTGAATAAAATATCA 195231

Qy         61 ArgSerGln------LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHis 78
              |||:::|||      :::           :::   |||      :::   |||
Db     195230 AGAGCCCAGTGCTGGCGACAACAGAGATCACTCAGCTACTGCACCCAAATATCTCCCTCT 195171

Qy         79 AlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
                 :::|||||||||           ::::::||||||   :::              |||
Db     195170 TTCCTCCTTACACCAGCT---------ATTTCCATGAAGTCTACA---------GTATCT 195129

Qy         99 SerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGly 118
              :::   ||||||   |||:::                                  |||
Db     195128 GCTGGTTCAAGCTCTGCACTT----------------------------ACGTACAAT 195099

Qy        119 LeuGluCysCysLeuLeuTyrLeu 126
              |||:::   ||||||   ||||||
Db     195098 TTACAGAAGTGCTTAGCATATCTT 195075

RESULT 9
AF291597
LOCUS       AF291597               9833 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Japanese 18 22q11.2 noncoding genomic
            sequence.
ACCESSION   AF291597
VERSION     AF291597.1  GI:10953199
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Japanese 18"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9833
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              8.07          Length:         9833
Score:                  95.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.69%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-84 (1-135) x AF291597 (1-9833)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGRTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT------------------------------------------ 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                   :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109
                 |||                                        |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118
                 :::   :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                         |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::      |||:::      |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 10
AF291608
LOCUS       AF291608               9432 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Berg 26 22q11.2 noncoding genomic sequence.
```

```
ACCESSION   AF291608
VERSION     AF291608.1  GI:10953210
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9432)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9432)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Berg 26"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9432
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              9.95          Length:         9432
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-84 (1-135) x AF291608 (1-9432)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                |||   |||:::|||                                   :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109
              |||                                           |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
```

```
                                                                        |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy         120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
               :::      |||:::      |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 11
AF291631
LOCUS       AF291631               9817 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate South African MeJe 22q11.2 noncoding genomic
            sequence.
ACCESSION   AF291631
VERSION     AF291631.1  GI:10953233
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9817)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9817)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9817
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="South African MeJe"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9817
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:         9817
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-84 (1-135) x AF291631 (1-9817)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||||:::
Db       2939 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 2998

Qy         24 LysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||||
Db       2999 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3034

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3035 GTAGATCTGGGATTT---------------------------------------------- 3049

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                |||   |||:::|||                                   :::|||
Db       3050 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3082

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
```

```
Db       3083 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3142

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109

Db       3143 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3202

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118

Db       3203 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3262

Qy        119 ------------------------------------------------------------Leu 119

Db       3263 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3322

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132

Db       3323 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3361
```

RESULT 12
AF291632
LOCUS AF291632 9817 bp DNA linear PRI 24-OCT-2000
DEFINITION Homo sapiens isolate South African Kung 22q11.2 noncoding genomic sequence.
ACCESSION AF291632
VERSION AF291632.1 GI:10953234
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 9817)
AUTHORS Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E., Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
PUBMED 11005839
REFERENCE 2 (bases 1 to 9817)
AUTHORS Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E., Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas at Houston Health Science Center, 6901 Bertner Ave., Houston, TX 77030, USA
FEATURES Location/Qualifiers
source 1. .9817
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="South African Kung"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature 1. .9817
/note="noncoding region"
ORIGIN

Alignment Scores:

Pred. No.:	10.4	Length:	9817
Score:	94.00	Matches:	42
Percent Similarity:	34.10%	Conservative:	17
Best Local Similarity:	24.28%	Mismatches:	38
Query Match:	13.54%	Indels:	76
DB:	9	Gaps:	5

US-09-974-546C-84 (1-135) x AF291632 (1-9817)

```
Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23

Db       2939 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 2998

Qy         24 LysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAlaProGluThr 43

Db       2999 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3034

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63

Db       3035 GTAGATCTGGGATTT---------------------------------------------- 3049

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83

Db       3050 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3082

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103

Db       3083 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3142

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109

Db       3143 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3202

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118

Db       3203 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3262

Qy        119 ------------------------------------------------------------Leu 119

Db       3263 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3322

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132

Db       3323 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3361
```

RESULT 13
AF291644
LOCUS AF291644 9830 bp DNA linear PRI 24-OCT-2000
DEFINITION Homo sapiens isolate Nigerian 9 22q11.2 noncoding genomic sequence.
ACCESSION AF291644
VERSION AF291644.1 GI:10953246
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 9830)
AUTHORS Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E., Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
PUBMED 11005839
REFERENCE 2 (bases 1 to 9830)
AUTHORS Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E., Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas at Houston Health Science Center, 6901 Bertner Ave., Houston, TX 77030, USA
FEATURES Location/Qualifiers
source 1. .9830
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="Nigerian 9"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature 1. .9830
/note="noncoding region"
ORIGIN

Alignment Scores:

Pred. No.:	10.4	Length:	9830
Score:	94.00	Matches:	42
Percent Similarity:	34.10%	Conservative:	17

```
Best Local Similarity:  24.28%          Mismatches:     38
Query Match:            13.54%          Indels:         76
DB:                     9               Gaps:           5

US-09-974-546C-84 (1-135) x AF291644 (1-9830)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT----------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                   :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle----------------------------------------LeuPheThrLeu 109
                |||                                        |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                        |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::     |||:::     |||     :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 14
AF291587
LOCUS       AF291587               9833 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Chinese 1 22q11.2 noncoding genomic sequence.
ACCESSION   AF291587
VERSION     AF291587.1  GI:10953189
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Chinese 1"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9833
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4            Length:         9833
Score:                  94.00           Matches:        42
Percent Similarity:     34.10%          Conservative:   17
Best Local Similarity:  24.28%          Mismatches:     38
Query Match:            13.54%          Indels:         76
DB:                     9               Gaps:           5

US-09-974-546C-84 (1-135) x AF291587 (1-9833)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT----------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                   :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle----------------------------------------LeuPheThrLeu 109
                |||                                        |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                        |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::     |||:::     |||     :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 15
AF291588
LOCUS       AF291588               9833 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Chinese 2 22q11.2 noncoding genomic sequence.
ACCESSION   AF291588
VERSION     AF291588.1  GI:10953190
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9833)
```

```
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Chinese 2"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9833
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:         9833
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-84 (1-135) x AF291588 (1-9833)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                  :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle----------------------------------------LeuPheThrLeu 109
               |||                                        |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly-------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ------------------------------------------------------------Leu 119
                                                                          |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::      |||:::     |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377


Search completed: September  9, 2005, 14:44:30
Job time : 2501 secs
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:02:29 ; Search time 309 Seconds
(without alignments)
2586.295 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174501_20885/app_query.fasta_1.654
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_468_@runat_07092005_174501_20885 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	2087	4	AAS04000	Aas04000 Biomarker
2	694	100.0	2088	3	AAZ87583	Aaz87583 Prostate
3	694	100.0	2505	4	AAS04001	Aas04001 Biomarker
4	694	100.0	2506	3	AAZ87584	Aaz87584 Prostate
5	90	13.0	1113	13	ADT45647	Adt45647 Bacterial
6	90	13.0	1146	13	ADS46591	Ads46591 Bacterial
7	89.5	12.9	4019	4	ABL27798	Abl27798 Drosophil
c 8	83.5	12.0	532	3	AAC38626	Aac38626 Arabidops
c 9	83	12.0	146547	8	ABZ80817	Abz80817 Human pho
c 10	82	11.8	24601	2	AAX13160	Aax13160 Enterococ
c 11	82	11.8	24601	6	ABS98955	Abs98955 Enterococ
c 12	81.5	11.7	133462	13	ABD32622	Abd32622 Mouse can
13	81	11.7	381	8	ABX44573	Abx44573 Bovine ES
c 14	81	11.7	110000	12	ADN46845_11	Continuation (12 o
15	81	11.7	110000	12	ADN47591_09	Continuation (10 o
c 16	81	11.7	110000	12	ADN46123_11	Continuation (12 o
17	81	11.7	110000	12	ADN47209_09	Continuation (10 o
c 18	81	11.7	110000	12	ADN46464_11	Continuation (12 o
19	81	11.7	110000	12	ADN47960_09	Continuation (10 o
c 20	80.5	11.6	1093	2	AAT95761	Aat95761 Arabidops
c 21	80.5	11.6	1093	3	AAC65293	Aac65293 Arabidops
22	80	11.5	1693	12	ADQ24476	Adq24476 Human sof
c 23	79.5	11.5	262	7	ADS31066	Ads31066 Human gen
24	79.5	11.5	585	5	ABV54205	Abv54205 Human pro
c 25	79.5	11.5	1464	3	AAC47238	Aac47238 Arabidops
26	79.5	11.5	6196	2	AAV13168	Aav13168 Complete
27	79.5	11.5	6243	2	AAV13841	Aav13841 Complete
28	79.5	11.5	6503	2	AAV13169	Aav13169 Complete
29	79.5	11.5	7379	2	AAV13176	Aav13176 Complete
30	79.5	11.5	8618	2	AAV18741	Aav18741 Complete
31	79.5	11.5	8792	2	AAV18745	Aav18745 Complete
c 32	78.5	11.3	386	8	ABX45278	Abx45278 Bovine ES
33	78.5	11.3	398	4	AAI89039	Aai89039 Human pol
34	78.5	11.3	4065	12	ADQ65017	Adq65017 Novel hum
c 35	78.5	11.3	26496	5	ABA19696	Aba19696 Human ner
c 36	78	11.2	3328	4	ABL18311	Abl18311 Drosophil
c 37	77.5	11.2	389	8	ABX39152	Abx39152 Bovine ES
38	77.5	11.2	567	10	ABX56974	Abx56974 Arabidops
c 39	77.5	11.2	2411	11	ADM02309	Adm02309 Human cDN
40	77.5	11.2	2437	12	ADP22463	Adp22463 Sea-squir
c 41	77.5	11.2	2661	4	ABL26351	Abl26351 Drosophil
42	77.5	11.2	3435	4	ABL02615	Abl02615 Drosophil
c 43	77.5	11.2	3958	4	ABL07870	Abl07870 Drosophil
c 44	77.5	11.2	4122	4	ABL07868	Abl07868 Drosophil
45	77.5	11.2	6010	10	ADF50278	Adf50278 Human PFM

ALIGNMENTS

RESULT 1
AAS04000
ID AAS04000 standard; cDNA; 2087 BP.
XX
AC AAS04000;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
XX
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 99. .506
FT /*tag= a
FT /product= "Prostate cancer marker protein"
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
XX
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.

XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri R;
XX
DR WPI; 2001-289849/30.
DR P-PSDB; AAU02174.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS Claim 2; Col 117-121; 78pp; English.
XX
CC The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.6e-76	Length:	2087
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-974-546C-84 (1-135) x AAS04000 (1-2087)

```
Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
```

RESULT 2
AAZ87583
ID AAZ87583 standard; DNA; 2088 BP.
XX
AC AAZ87583;

XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59295.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Claim 1; Page 182-183; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	2.6e-76	Length:	2088
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-974-546C-84 (1-135) x AAZ87583 (1-2088)

```
Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
```

```
Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
```

RESULT 3
AAS04001
ID AAS04001 standard; cDNA; 2505 BP.
XX
AC AAS04001;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.
XX
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 99. .506
FT /*tag= a
FT /product= "Prostate cancer marker protein"
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
XX
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri R;
XX
DR WPI; 2001-289849/30.
DR P-PSDB; AAU02175.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS Claim 2; Col 121-125; 78pp; English.
XX
CC The sequence represents nucleic acid biomarker UC band 28 #3, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX

SQ Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.39e-76	Length:	2505
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-974-546C-84 (1-135) x AAS04001 (1-2505)

```
Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
```

RESULT 4
AAZ87584
ID AAZ87584 standard; DNA; 2506 BP.
XX
AC AAZ87584;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 splice variant.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59296.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.

```
XX
PS   Claim 1; Page 184-186; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              3.39e-76        Length:         2506
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-974-546C-84 (1-135) x AAZ87584 (1-2506)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 5
ADT45647
ID   ADT45647 standard; cDNA; 1113 BP.
XX
AC   ADT45647;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polynucleotide #20398.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 44085; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1113 BP; 379 A; 204 C; 242 G; 288 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.53            Length:         1113
Score:                  90.00           Matches:        34
Percent Similarity:     44.88%          Conservative:   23
Best Local Similarity:  26.77%          Mismatches:     38
Query Match:            12.97%          Indels:         32
DB:                     13              Gaps:           7

US-09-974-546C-84 (1-135) x ADT45647 (1-1113)

Qy         13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu--------- 29
              |||||||||:::         ::::::::::|||:::      |||   :::
Db         61 ATGCACAACTTA------GCCATAAAGCTGAGAGAAAGAGGGCATGAAGTGGGAATAGTC 114
```

```
Qy          30 ---SerAsnPhePro---------------------------------------------- 33
                  :::|||   |||
Db         115 ACAAACAACAGGCCCACGGGAAAGGAAGAAGAGCTTAAGAGATATGGAATAGAGCTCATA 174

Qy          34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
               ::::::||||||||::::::|||                              :::
Db         175 AAGATCCCAGGAATTATAAGTCCT----------------------TTTTAGATGTA 210

Qy          54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle---AspTyrSer 72
                           :::|||   |||::::::      ||||||:::   |||:::
Db         211 AATTTAACCTATGGATTGAAATCCTCAGAAGAGCTCAACGAATTCTTGAAAGACTTCGAC 270

Qy          73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
               |||      ||||||||||||      |||||||||                     |||
Db         271 ATAATTCATTCTCATCATGCA---TTCACACCTCTCTCCTTAAAGGCTTTAAAAGCTGGA 327

Qy          93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                  :::|||   |||   :::   |||::::::   :::|||   |||
Db         328 AAGAATATGGAAAAGGGAACCCTGCTAACAACTCACAGCATTTCCTTTGCCCATGAATCA 387

Qy         112 GlnLeuThrGlnThrLeuGly 118
               :::|||      |||||||||
Db         388 AAGCTCTGGGATACTTTGGGG 408
```

RESULT 6
ADS46591
ID ADS46591 standard; cDNA; 1146 BP.
XX
AC ADS46591;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #1334.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 25021; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1146 BP; 388 A; 207 C; 253 G; 298 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.553	Length:	1146
Score:	90.00	Matches:	34
Percent Similarity:	44.88%	Conservative:	23
Best Local Similarity:	26.77%	Mismatches:	38
Query Match:	12.97%	Indels:	32
DB:	13	Gaps:	7

US-09-974-546C-84 (1-135) x ADS46591 (1-1146)

```
Qy          13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu--------- 29
               |||||||||:::         :::::::::|||:::      |||   :::
Db          70 ATGCACAACTTA------GCCATAAAGCTGAGAGAAAGAGGGCATGAAGTGGGAATAGTC 123

Qy          30 ---SerAsnPhePro---------------------------------------------- 33
                  :::|||   |||
Db         124 ACAAACAACAGGCCCACGGGAAAGGAAGAAGAGCTTAAGAGATATGGAATAGAGCTCATA 183

Qy          34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
               ::::::||||||||::::::|||                              :::
Db         184 AAGATCCCAGGAATTATAAGTCCT----------------------TTTTAGATGTA 219

Qy          54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle---AspTyrSer 72
                           :::|||   |||::::::      ||||||:::   |||:::
Db         220 AATTTAACCTATGGATTGAAATCCTCAGAAGAGCTCAACGAATTCTTGAAAGACTTCGAC 279

Qy          73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
               |||      ||||||||||||      |||||||||                     |||
Db         280 ATAATTCATTCTCATCATGCA---TTCACACCTCTCTCCTTAAAGGCTTTAAAAGCTGGA 336

Qy          93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                  :::|||   |||   :::   |||::::::   :::|||   |||
Db         337 AAGAATATGGAAAAGGGAACCCTGCTAACAACTCACAGCATTTCCTTTGCCCATGAATCA 396

Qy         112 GlnLeuThrGlnThrLeuGly 118
               :::|||      |||||||||
Db         397 AAGCTCTGGGATACTTTGGGG 417
```

RESULT 7
ABL27798
ID ABL27798 standard; DNA; 4019 BP.
XX
AC ABL27798;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34867.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX

OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 34867; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4019 BP; 1158 A; 966 C; 927 G; 968 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.93	Length:	4019
Score:	89.50	Matches:	35
Percent Similarity:	39.84%	Conservative:	16
Best Local Similarity:	27.34%	Mismatches:	40
Query Match:	12.90%	Indels:	37
DB:	4	Gaps:	5

US-09-974-546C-84 (1-135) x ABL27798 (1-4019)

```
Qy         18 HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGly 37
              |||   |||:::   ||||||::::::||||||      ||||||   ||:::|||
Db       1389 CATAGACTAAAAGCACGAAAGGTGCAACACCGTAATTATAATTTTAGACGTGTTCCAGCG 1448

Qy         38 IleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGlu 57
                  |||   |||                 ::::::      :::::::|||
Db       1449 CATAGGCAGCCCATCACA---------TCCACATCGCATCAAAGAACGAAAAGAAAAGCA 1499

Qy         58 LysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHis 77
              :::       |||      |||   |||      |||                    |||
Db       1500 CGTCAGAGCAGGGAAACTAAGTGTACTTCGAAGATA--------------------CAT 1538

Qy         78 HisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCys 97
                 |||      ||||||                                  :::|||
Db       1539 TGTACAATTTACATGCCGCTG--------------------------------AGGTGC 1565

Qy         98 SerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeu 117
              :::|||   :::            |||   :::|||:::|||||||||   :::
Db       1566 ACCTCGGAAAACCTTAATTGGTTTTTAGCCTCATTAAGCTTGCAGCTAGAAAGAATGTGG 1625

Qy        118 GlyLeuGlu---------------------------CysCys----------------- 122
              |||   |||                          ||||||
Db       1626 GGTCAAGAATAGCCATCAAATATTGCTCATACGACGTGTTGTACATGCGATTAGTGTAGG 1685

Qy        123 ---LeuLeuTyrLeuSerLysThr 129
                  :::||||||      |||
Db       1686 GAAGCGATGTATCTCCCTCTTACA 1709
```

RESULT 8
AAC38626/c
ID AAC38626 standard; DNA; 532 BP.
XX
AC AAC38626;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21658.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	1.19	Length:	532
Score:	83.50	Matches:	39
Percent Similarity:	41.48%	Conservative:	17
Best Local Similarity:	28.89%	Mismatches:	52
Query Match:	12.03%	Indels:	27
DB:	3	Gaps:	7

US-09-974-546C-84 (1-135) x AAC38626 (1-532)

```
Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              |||  |||           :::  :::      |||           |||
Db        445 TACAGCGACAGATGTGTTTTTGTTGTGCTCTTCTTTATCCACTCAAGTAACCATCATCAC 386

Qy         30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
```

```
                |||   |||   |||      :::      |||                    |||
Db        385 TATAACTTACCCAATCTAGTTTGGGTTTGGCTTCCACACGGAACA-----------TGT 338

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe--- 68
                           |||:::|||:::||||||      |||::::::|||
Db        337 AGCCATGGGTCGTCTCTTCCAGAAGAGGTAGAGCGCTCCTTCATGGCCTCCCAGTTTGTA 278

Qy         69 ---------IleAspTyrSerIleGluGlnSerHisHisAla------IleLeuThr--- 82
                   |||   :::||||||      :::   :::|||      :::||||||
Db        277 GTTTTTGTCATAACTTTCAGCATCTCACTCGCTGCAAACGCTGCAGATGTGCTGACCGGC 218

Qy         83 ------------ProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
                        |||            ||||||:::      |||      |||
Db        217 CAACCCGTGAACCCAGCCATCATCATTCTGACCCTCCATTTCCCCAGTACCTCGTGTCTC 158

Qy         99 SerLeu---------SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGln 115
              |||            ||||||   ||||||:::||||||:::
Db        157 TCTACTCTCTCTGACTCCTCACATGCTATCATGTTCACTATG----------------- 116

Qy        116 ThrLeuGlyLeuGluCysCys---------LeuLeuTyrLeuSer 127
              :::|||      :::||||||         :::|||   ||||||
Db        115 TCTCTTGCTACACAGTGTTGTTCTGCGCTGATTCTCTGCTTATCA 71

RESULT 9
ABZ80817/c
ID   ABZ80817 standard; DNA; 146547 BP.
XX
AC   ABZ80817;
XX
DT   13-JUN-2003  (first entry)
XX
DE   Human phospholipase C gamma 1 polymorphism G329u1.
XX
KW   human; single nucleotide polymorphism; SNP; phospholipase C gamma 1;
KW   PLCG1; gene; vascular disease; plasminogen activator inhibitor type 2;
KW   PAI-2; diagnosis; atherosclerosis; coronary artery disease; ischemia;
KW   myocardial infarction; stroke; thromboembolism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(64001,c)
FT                   /*tag=  a
XX
PN   WO2003007801-A2.
XX
PD   30-JAN-2003.
XX
PF   19-JUL-2002; 2002WO-US023041.
XX
PR   20-JUL-2001; 2001US-0306941P.
PR   28-AUG-2001; 2001US-0315572P.
PR   05-OCT-2001; 2001US-0327488P.
PR   14-DEC-2001; 2001US-00017128.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J;
XX
DR   WPI; 2003-278312/27.
XX
PT   Diagnosing and prognosing vascular disease, e.g. coronary artery disease
PT   and myocardial infarction, based on the detection of polymorphisms in the
PT   phospholipase C gamma 1 and plasminogen activator inhibitor type 2 genes
PT   and polypeptides.
XX
PS   Claim 44; Page 204-245; 295pp; English.
XX
CC   The invention relates to methods, agents and apparatus for diagnosing and
CC   prognosing vascular diseases based on the detection of polymorphisms in
CC   the phospholipase C gamma 1 (PLCG1) and plasminogen activator inhibitor
CC   type 2 (PAI-2) genes and polypeptides. In a preferred claim of the
CC   invention the polymorphism is a C to T nucleotide change at position
CC   64001 of the PLCG1 gene (reference sequence GI 11345540) causing an amino
CC   acid change from Threonine to Isoleucine at position 813 of the protein.
CC   This sequence represents the region of the PLCG1 gene with the T form of
CC   the polymorphism. The methods, agents and apparatus is used for
CC   diagnosing and prognosing vascular diseases e.g. atherosclerosis,
CC   coronary artery disease (especially), myocardial infarction (especially),
CC   ischemia, stroke, peripheral vascular diseases, venous thromboembolism
CC   and pulmonary embolism
XX
SQ   Sequence 146547 BP; 39552 A; 32625 C; 33339 G; 41031 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.69e+03        Length:         146547
Score:                  83.00           Matches:        19
Percent Similarity:     65.71%          Conservative:   4
Best Local Similarity:  54.29%          Mismatches:     10
Query Match:            11.96%          Indels:         2
DB:                     8               Gaps:           1

US-09-974-546C-84 (1-135) x ABZ80817 (1-146547)

Qy         16 IleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeu 35
              ||||||   |||   |||::::::||||||||||||||||
Db      29020 ATTATCCCCATTTTCCAGATGAAAAAACTGAGGCACAGAGCTAAAAGGCAGGATTCAAAC 28961

Qy         36 ProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyr 50
              ||||||   ||||||||||||:::||||||      ||||||:::
Db      28960 CCAGGTAGTCTGGCCCCAGAGTCAGTGCTG------TTTTGCTTC 28922

RESULT 10
AAX13160/c
ID   AAX13160 standard; DNA; 24601 BP.
XX
AC   AAX13160;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:223.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US008985.
XX
PR   06-MAY-1997;   97US-0044031P.
PR   16-MAY-1997;   97US-0046655P.
PR   14-NOV-1997;   97US-0066009P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA,  Dillon PJ,  Barash SC;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides -
PT   used to develop products for the detection of Enterococcus and for use in
PT   vaccines for prevention or attenuation of Enterococcus infection.
XX
PS   Claim 1; Page 1125-1137; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence of
CC   Enterococcus faecalis in samples. They can also be used for diagnosing
```

```
CC   Enterococcal infection in an animal and monitoring progression of
CC   disease, and for identifying agents which can be used to modulate the
CC   growth or pathogenicity of Enterococcus faecalis, or another related
CC   organism, in vivo or in vitro. In particular the polypeptides encoded by
CC   the Enterococcus faecalis nucleotide sequences can be used in vaccines to
CC   prevent or attenuate an Enterococcal infection
XX
SQ   Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
Pred. No.:              472          Length:        24601
Score:                  82.00        Matches:       42
Percent Similarity:     47.10%       Conservative:  23
Best Local Similarity:  30.43%       Mismatches:    51
Query Match:            11.82%       Indels:        22
DB:                     2            Gaps:          8

US-09-974-546C-84 (1-135) x AAX13160 (1-24601)

Qy          7 AsnGlnLysTyrGluAsp--------------------MetHisAsnIleIle------ 17
              |||||||||   :::                        |||::::::|||
Db       2571 AACCAAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTTCACGACTTAATAAAACAC 2512

Qy         18 ---------------HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
                            |||   |||:::   |||   ::::::
Db       2511 GCTTAAAATAAGGCACATTCACTAAAAGAAAGGCCAGTAAAATAATTAAGCCCACCCGTT 2452

Qy         33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
              |||   |||      |||:::   |||         |||||||||      |||   :::
Db       2451 CCATCATTAAAATAAATAATTCGACCACACCGGCACCTCCTTCCTATTAACTATACTATA 2392

Qy         53 PheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSer 72
                 ::::::|||||||||   :::   |||||||||        ||||||   :::|||
Db       2391 AATCAAAGAAAAGAAAAAAATGAAGCTTCTCAAAAA---AGAGGCTTCATT---TTTTCT 2338

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              |||         :::      ||||||   ||||||:::      |||
Db       2337 ATT---CATTAAAACGCAGGAATATTGGCGCCTTTAGAATTTTTCTTGATGTATTCAGCA 2281

Qy         93 SerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThr------ 110
              :::|||            :::   |||   :::   ::::::|||:::|||:::
Db       2280 ACTTCTTTTGTTGTCATCGCT---TCTTTTAAAACCTTGATTTTTTCGCTATCTTTTTCG 2224

Qy        111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
                 |||::::::|||   |||           |||   |||      ||||||
Db       2223 CCTTTGCGTGTCACA---ACTTGCAATGCGTAATGTTTACTTGCTTCTTTATCT 2173

RESULT 11
ABS98955/c
ID   ABS98955 standard; DNA; 24601 BP.
XX
AC   ABS98955;
XX
DT   18-DEC-2002  (first entry)
XX
DE   Enterococcus faecalis contig sequence #223.
XX
KW   Computer readable medium; Enterococcus faecalis; microbe; growth;
KW   pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
KW   therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
KW   biochip technology; antibacterial; modulator of nucleic acid expression;
KW   contig; ds.
XX
OS   Enterococcus faecalis.
XX
PN   US2002120116-A1.
XX
PD   29-AUG-2002.
XX
PF   04-MAY-1998;   98US-00070927.
XX
PR   04-MAY-1998;   98US-00070927.
XX
PA   (KUNS/) KUNSCH C A.
PA   (DILL/) DILLON P J.
PA   (BARA/) BARASH S.
XX
PI   Kunsch CA,  Dillon PJ,  Barash S;
XX
DR   WPI; 2002-750065/81.
XX
PT   Computer readable medium having recorded on it a Enterococcus faecalis
PT   nucleotide sequence useful for detecting diseases related to Enterococcus
PT   infections in animals.
XX
PS   Claim 1; Page; 119pp; English.
XX
CC   The present invention relates to a new computer readable medium with an
CC   Enterococcus faecalis nucleotide sequence. The invention is useful to
CC   diagnose the presence of E.faecalis in a sample or determining the
CC   presence of a specific microbe in a sample. The invention is also useful
CC   for modulating the growth or pathogenicity of E.faecalis, in a vaccine to
CC   confer resistance to Enterococcal infection, for commercial, therapeutic
CC   and industrial purposes, and for fermenting a particular sugar source or
CC   to produce a particular metabolite. The invention is useful for detecting
CC   diseases related to Enterococcus infections in animals, and for detecting
CC   E.faecalis using biochip technology. The present nucleic acid sequence
CC   represents an Enterococcus faecalis contig DNA sequence of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at http.sequdata.uspto.gov
XX
SQ   Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
Pred. No.:              472          Length:        24601
Score:                  82.00        Matches:       42
Percent Similarity:     47.10%       Conservative:  23
Best Local Similarity:  30.43%       Mismatches:    51
Query Match:            11.82%       Indels:        22
DB:                     6            Gaps:          8

US-09-974-546C-84 (1-135) x ABS98955 (1-24601)

Qy          7 AsnGlnLysTyrGluAsp--------------------MetHisAsnIleIle------ 17
              |||||||||   :::                        |||::::::|||
Db       2571 AACCAAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTTCACGACTTAATAAAACAC 2512

Qy         18 ---------------HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
                            |||   |||:::   |||   ::::::
Db       2511 GCTTAAAATAAGGCACATTCACTAAAAGAAAGGCCAGTAAAATAATTAAGCCCACCCGTT 2452

Qy         33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
              |||   |||      |||:::   |||         |||||||||      |||   :::
Db       2451 CCATCATTAAAATAAATAATTCGACCACACCGGCACCTCCTTCCTATTAACTATACTATA 2392

Qy         53 PheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSer 72
                 ::::::|||||||||   :::   |||||||||        ||||||   :::|||
Db       2391 AATCAAAGAAAAGAAAAAAATGAAGCTTCTCAAAAA---AGAGGCTTCATT---TTTTCT 2338

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              |||         :::      ||||||   ||||||:::      |||
Db       2337 ATT---CATTAAAACGCAGGAATATTGGCGCCTTTAGAATTTTTCTTGATGTATTCAGCA 2281

Qy         93 SerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThr------ 110
              :::|||            :::   |||   :::   ::::::|||:::|||:::
Db       2280 ACTTCTTTTGTTGTCATCGCT---TCTTTTAAAACCTTGATTTTTTCGCTATCTTTTTCG 2224

Qy        111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
                 |||::::::|||   |||           |||   |||      ||||||
Db       2223 CCTTTGCGTGTCACA---ACTTGCAATGCGTAATGTTTACTTGCTTCTTTATCT 2173

RESULT 12
ABD32622/c
ID   ABD32622 standard; DNA; 133462 BP.
XX
```

```
AC   ABD32622;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Mouse cancer-associated genomic DNA MD13-011.
XX
KW   Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Mus musculus.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   disclosure; seqid 140; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a mouse CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 133462 BP; 39091 A; 24305 C; 25093 G; 41557 T; 0 U; 3416 Other;

Alignment Scores:
Pred. No.:              6.32e+03        Length:         133462
```

```
Score:                  81.50           Matches:        37
Percent Similarity:     42.86%          Conservative:   17
Best Local Similarity:  29.37%          Mismatches:     33
Query Match:            11.74%          Indels:         39
DB:                     13              Gaps:           8

US-09-974-546C-84 (1-135) x ABD32622 (1-133462)

Qy         17 IleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuPro 36
              :::|||   |||          :::|||        :::||||||   :::
Db       9988 CTACATGCACTGCTAAATGGTGATGTCAGATTCGGTTTTACTAACTTCTCCAAA------ 9935

Qy         37 GlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLys 56
                           ||||||      |||   |||   |||:::         |||
Db       9934 --------------AATACTGTAAAGGGTCCTCTTTGCAGGAAGTTA---------AAA 9899

Qy         57 GluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSer 76
              :::|||   |||||||||   |||   :::             |||   :::      |||
Db       9898 AAAAAAAAAAAAAAGGTCGTTCAAAGTCAGTGCAATGGCTTCTTACTTCCTTAATTCAAGT 9839

Qy         77 HisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLys 96
              ::::::|||:::                                |||||||||   :::
Db       9838 AATTATGCTGTT----------------------------AGCGGTTCGTCT---CAA 9812

Qy         97 CysSerSerLeuSerSerGlu-----------------AlaIleLeuPheThrLeuThr 110
                 ||||||         :::                :::||||||   :::
Db       9811 TTTTCCTCTAAAGGAAAACAGTTTTAAACTATATAAATATTGCTCCTTTTTAATATCTTT 9752

Qy        111 LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysThrIle 130
              |||:::                          ||||||   ||||||||||||
Db       9751 TTAAGG----------------------------AAGCTTTAT---TCAAAAACAATA 9725

Qy        131 HisPro---GlnIleIle 135
              ||||||   |||:::|||
Db       9724 CATCCCTCCCAAGTCATC 9707
```

```
RESULT 13
ABX44573
ID   ABX44573 standard; cDNA; 381 BP.
XX
AC   ABX44573;
XX
DT   21-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #9738.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-00960352.
XX
PR   12-JAN-1999;   99US-0115707P.
PR   11-JAN-2000; 2000US-00480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
```

```
XX
PS   Claim 2; SEQ ID NO 9738; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3' non- translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 381 BP; 104 A; 84 C; 82 G; 111 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.51          Length:          381
Score:                    81.00         Matches:         24
Percent Similarity:       42.55%        Conservative:    16
Best Local Similarity:    25.53%        Mismatches:      30
Query Match:              11.67%        Indels:          24
DB:                       8             Gaps:            4

US-09-974-546C-84 (1-135) x ABX44573 (1-381)

Qy         24 LysLeuArgHisArgLeuSerAsnPhePro------ArgLeuProGly------------ 37
              |||:::||||||:::::          |||      |||||||||
Db         96 AAGATAAGACACCAGATTCTCAAACTGCCAACTAGTCGTCTGCCCAGTAGCTGACGACGG 155

Qy         38 ------------------------IleLeuAlaProGluThrValLeuLeuProPheCys 49
                                     :::   |||       :::   |||
Db        156 AAAGGTTTGCCAACCCGTTCGCCCGTACACCTGCCCCGTTTCACGATGCAACCTGCTTCT 215

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
                 |||   :::   :::::::::   |||                      ||||||
Db        216 GCAAAGTGGTACGATCGGAGGGACTATGTC-----------------------TTCATT 251

Qy         70 AspTyrSerIleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThr 89
              ::::::   :::|||   |||   |||      :::   ||||||::::::   ||||||
Db        252 GAATTTTGTGTCGAAGACAGTCAACATGTTAATGTAGATCCTTTAAAATCGAAACTTACA 311

Qy         90 MetLys------GlySerSerMetLysCysSerSerLeuSer 101
                    |||   |||      |||      |||:::
Db        312 TTCAGTTGTCTTGGAGGAAGTGATAATTGTAAACATTTAAAT 353

RESULT 14
ADN46845_11/c
Continuation (12 of 21) of ADN46845 from base 1100001 (Thermococcus kodakaraensis KOD1 g
WP Sequence split into 21 fragments  LOCUS ADN46845 Accession Adn46845
WP     Fragment Name      Begin     End
WP     ADN46845_00            1    110000
WP     ADN46845_01       100001    210000
WP     ADN46845_02       200001    310000
WP     ADN46845_03       300001    410000
WP     ADN46845_04       400001    510000
WP     ADN46845_05       500001    610000
WP     ADN46845_06       600001    710000
WP     ADN46845_07       700001    810000
WP     ADN46845_08       800001    910000
WP     ADN46845_09       900001   1010000
WP     ADN46845_10      1000001   1110000
WP     ADN46845_11      1100001   1210000
WP     ADN46845_12      1200001   1310000
WP     ADN46845_13      1300001   1410000
WP     ADN46845_14      1400001   1510000
WP     ADN46845_15      1500001   1610000
WP     ADN46845_16      1600001   1710000
WP     ADN46845_17      1700001   1810000
WP     ADN46845_18      1800001   1910000
WP     ADN46845_19      1900001   2010000
WP     ADN46845_20      2000001   2089378

Alignment Scores:
Pred. No.:                5.52e+03      Length:          110000
Score:                    81.00         Matches:         31
Percent Similarity:       44.36%        Conservative:    28
Best Local Similarity:    23.31%        Mismatches:      42
Query Match:              11.67%        Indels:          32
DB:                       12            Gaps:            7

US-09-974-546C-84 (1-135) x ADN46845_11 (1-110000)

Qy         13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer------ 30
              ||||||   :::         :::   ::::::||||||   |||   :::|||
Db      68597 ATGCACCAGCTT------GCAATCTATCTGAAAAAGCTCGGCCACGACGTTTCAATAGTG 68544

Qy         31 ------------------------------------------------AsnPhePro 33

Db      68543 ACTAATGATCTAAAAACGGGAAAGGAAAAAGAGCTTGAAGAGCTGGGGGTAGGATTGGTT 68484

Qy         34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
              ::::::||||||:::::::::|||      :::   :::         |||
Db      68483 AAAGTCCCGGGTGTCATCAGTCCG---GTGCTCGGGATAAACATAACCTAC--------- 68436

Qy         54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72
                        :::|||   :::::::::      ||||||   :::|||:::
Db      68435 -----------GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTTCTCGTGGATTTTGAT 68388

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              :::      :::|||||||||      |||||||||                   |||
Db      68387 GTTGTCCACGCCCACCACGCC---TTCACTCCACTCTCGCTGAAGGCGGTTAAGGCCGGG 68331

Qy         93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
              ::::::   |||   :::   |||::::::   :::|||   ||:::
Db      68330 AGAACTCTCGAAAAGGCAACACTCCTCACAACCCACAGCATATCCTTCTCCCACGAGTCT 68271

Qy        112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124
                 |||   :::   ||||||||||         ||||||
Db      68270 TCCCTATGGAAGGCCCTTGGGTTGACTTTTCCACTCCTC 68232

RESULT 15
ADN47591_09
Continuation (10 of 21) of ADN47591 from base 900001 (Thermococcus kodakaraensis KOD1 ger
WP Sequence split into 21 fragments  LOCUS ADN47591 Accession Adn47591
WP     Fragment Name      Begin     End
WP     ADN47591_00            1    110000
WP     ADN47591_01       100001    210000
WP     ADN47591_02       200001    310000
WP     ADN47591_03       300001    410000
WP     ADN47591_04       400001    510000
WP     ADN47591_05       500001    610000
WP     ADN47591_06       600001    710000
WP     ADN47591_07       700001    810000
WP     ADN47591_08       800001    910000
WP     ADN47591_09       900001   1010000
WP     ADN47591_10      1000001   1110000
WP     ADN47591_11      1100001   1210000
WP     ADN47591_12      1200001   1310000
WP     ADN47591_13      1300001   1410000
```

```
WP      ADN47591_14       1400001   1510000
WP      ADN47591_15       1500001   1610000
WP      ADN47591_16       1600001   1710000
WP      ADN47591_17       1700001   1810000
WP      ADN47591_18       1800001   1910000
WP      ADN47591_19       1900001   2010000
WP      ADN47591_20       2000001   2089378

Alignment Scores:
Pred. No.:              5.52e+03        Length:         110000
Score:                  81.00           Matches:        31
Percent Similarity:     44.36%          Conservative:   28
Best Local Similarity:  23.31%          Mismatches:     42
Query Match:            11.67%          Indels:         32
DB:                     12              Gaps:           7

US-09-974-546C-84 (1-135) x ADN47591_09 (1-110000)

Qy         13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer------ 30
              ||||||   :::         :::   ::::::||||||   |||   :::|||
Db      20781 ATGCACCAGCTT------GCAATCTATCTGAAAAAGCTCGGCCACGACGTTTCAATAGTG 20834

Qy         31 ----------------------------------------------------AsnPhePro 33

Db      20835 ACTAATGATCTAAAAACGGGAAAGGAAAAAGAGCTTGAAGAGCTGGGGGTAGGATTGGTT 20894

Qy         34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
              ::::::||||||:::::::::|||      :::   :::         |||
Db      20895 AAAGTCCCGGGTGTCATCAGTCCG---GTGCTCGGGATAAACATAACCTAC--------- 20942

Qy         54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72
                        :::|||   :::::::::      ||||||   :::|||:::
Db      20943 ------------GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTTCTCGTGGATTTTGAT 20990

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              :::      :::|||||||||       |||||||||                  |||
Db      20991 GTTGTCCACGCCCACCACGCC---TTCACTCCACTCTCGCTGAAGGCGGTTAAGGCCGGG 21047

Qy         93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                 ::::::   |||   :::   |||::::::   :::|||   |||:::
Db      21048 AGAACTCTCGAAAAGGCAACACTCCTCACAACCCACAGCATATCCTTCTCCCACGAGTCT 21107

Qy        112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124
                 |||   :::   |||||||||         ||||||
Db      21108 TCCCTATGGAAGGCCCTTGGGTTGACTTTTCCACTCCTC 21146


Search completed: September  9, 2005, 13:23:49
Job time : 392 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:12:05 ; Search time 104.5 Seconds
(without alignments)
2113.850 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174503_21006/app_query.fasta_1.654
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_77_@runat_07092005_174503_21006 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	694	100.0	2087	3	US-09-097-199-83	Sequence 83, Appl
	2	694	100.0	2505	3	US-09-097-199-85	Sequence 85, Appl
	3	694	100.0	2506	4	US-09-949-016-3311	Sequence 3311, Ap
	4	694	100.0	6507	4	US-09-949-016-15053	Sequence 15053, A
c	5	689	99.3	601	4	US-09-949-016-119365	Sequence 119365,
c	6	689	99.3	601	4	US-09-949-016-119366	Sequence 119366,
c	7	480	69.2	601	4	US-09-949-016-119364	Sequence 119364,
	8	83.5	12.0	2472	4	US-09-248-796A-148	Sequence 148, App
c	9	80.5	11.6	1093	3	US-09-186-276B-53	Sequence 53, Appl
c	10	80.5	11.6	1093	3	US-08-842-445-53	Sequence 53, Appl
c	11	80.5	11.6	1093	3	US-09-186-188B-53	Sequence 53, Appl
c	12	80.5	11.6	1093	4	US-09-265-585C-53	Sequence 53, Appl
c	13	80	11.5	148783	4	US-09-949-016-15729	Sequence 15729, A
c	14	79.5	11.5	262	4	US-09-573-080A-99	Sequence 99, Appl
	15	79.5	11.5	6196	3	US-08-675-566-5	Sequence 5, Appli
	16	79.5	11.5	6243	3	US-08-675-566-14	Sequence 14, Appl
	17	79.5	11.5	6503	3	US-08-675-566-6	Sequence 6, Appli
	18	79.5	11.5	7379	3	US-08-675-566-13	Sequence 13, Appl
	19	79.5	11.5	8618	3	US-08-675-566-21	Sequence 21, Appl
	20	79.5	11.5	8792	3	US-08-675-566-25	Sequence 25, Appl
	21	78	11.2	761	4	US-09-270-767-1826	Sequence 1826, Ap
	22	78	11.2	761	4	US-09-270-767-17108	Sequence 17108, A
	23	78	11.2	96878	4	US-09-949-016-12551	Sequence 12551, A
c	24	78	11.2	462589	4	US-09-949-016-12900	Sequence 12900, A
c	25	78	11.2	476044	4	US-09-949-016-12412	Sequence 12412, A
	26	77	11.1	1461	1	US-08-587-670A-1	Sequence 1, Appli
	27	77	11.1	1461	3	US-09-061-674-1	Sequence 1, Appli
	28	77	11.1	22372	4	US-09-949-016-17459	Sequence 17459, A
	29	76	11.0	367	3	US-09-328-111-446	Sequence 446, App
c	30	76	11.0	601	4	US-09-949-016-202994	Sequence 202994,
	31	76	11.0	42672	4	US-09-949-016-17253	Sequence 17253, A
	32	76	11.0	42672	4	US-09-949-016-17254	Sequence 17254, A
	33	76	11.0	66213	4	US-09-949-016-11803	Sequence 11803, A
	34	76	11.0	66213	4	US-09-949-016-16739	Sequence 16739, A
c	35	75.5	10.9	19008	4	US-09-949-016-12923	Sequence 12923, A
c	36	75.5	10.9	160759	4	US-09-949-016-16514	Sequence 16514, A
	37	75	10.8	64291	4	US-09-949-016-16278	Sequence 16278, A
	38	75	10.8	117410	4	US-09-949-016-12262	Sequence 12262, A
	39	74.5	10.7	3915	4	US-09-023-655-1104	Sequence 1104, Ap
c	40	74.5	10.7	5522	4	US-09-949-016-1259	Sequence 1259, Ap
c	41	74.5	10.7	5523	4	US-09-949-016-1014	Sequence 1014, Ap
	42	74.5	10.7	27120	4	US-09-949-016-16210	Sequence 16210, A
	43	74.5	10.7	35707	4	US-09-949-016-17120	Sequence 17120, A
c	44	74	10.7	601	4	US-09-949-016-157153	Sequence 157153,
c	45	74	10.7	3143	4	US-09-949-016-856	Sequence 856, App

ALIGNMENTS

RESULT 1
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:

```
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  83:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2087 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..503
US-09-097-199-83

Alignment Scores:
Pred. No.:              1.65e-85        Length:         2087
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-974-546C-84 (1-135) x US-09-097-199-83 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
US-09-097-199-85
; Sequence 85, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  85:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..503
US-09-097-199-85

Alignment Scores:
Pred. No.:              2.21e-85        Length:         2505
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-974-546C-84 (1-135) x US-09-097-199-85 (1-2505)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 3
US-09-949-016-3311
; Sequence 3311, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3311
;   LENGTH: 2506
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-3311

Alignment Scores:
Pred. No.:              2.21e-85      Length:         2506
Score:                  694.00        Matches:        135
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     4             Gaps:           0

US-09-974-546C-84 (1-135) x US-09-949-016-3311 (1-2506)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 156

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 216

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 276

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 336

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 396

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 456

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        457 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 501

RESULT 4
US-09-949-016-15053
; Sequence 15053, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15053
;   LENGTH: 6507
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15053

Alignment Scores:
Pred. No.:              9.96e-85      Length:         6507
Score:                  694.00        Matches:        135
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     4             Gaps:           0

US-09-974-546C-84 (1-135) x US-09-949-016-15053 (1-6507)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2097 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 2156

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2157 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 2216

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2217 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 2276

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2277 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 2336

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2337 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 2396

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2397 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 2456

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db       2457 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 2501

RESULT 5
US-09-949-016-119365/c
; Sequence 119365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119365
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119365

Alignment Scores:
Pred. No.:              1.14e-85      Length:         601
Score:                  689.00        Matches:        134
```

```
Percent Similarity:     99.26%          Conservative:  0
Best Local Similarity:  99.26%          Mismatches:    1
Query Match:            99.28%          Indels:        0
DB:                     4               Gaps:          0

US-09-974-546C-84 (1-135) x US-09-949-016-119365 (1-601)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 447

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 387

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 327

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        326 AGAAGTCAAAAGGCAACAGAGTTCAYTGATTATTCCATAGAACAGTCACACCATGCAATT 267

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 207

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 147

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        146 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 102

RESULT 6
US-09-949-016-119366/c
; Sequence 119366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119366
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119366

Alignment Scores:
Pred. No.:              1.14e-85        Length:        601
Score:                  689.00          Matches:       134
Percent Similarity:     99.26%          Conservative:  0
Best Local Similarity:  99.26%          Mismatches:    1
Query Match:            99.28%          Indels:        0
DB:                     4               Gaps:          0

US-09-974-546C-84 (1-135) x US-09-949-016-119366 (1-601)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 459

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 399

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 339

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              |||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db        338 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCAYAGAACAGTCACACCATGCAATT 279

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 219

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 159

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        158 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 114

RESULT 7
US-09-949-016-119364/c
; Sequence 119364, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119364
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119364

Alignment Scores:
Pred. No.:              9.72e-57        Length:        601
Score:                  480.00          Matches:       92
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            69.16%          Indels:        0
DB:                     4               Gaps:          0

US-09-974-546C-84 (1-135) x US-09-949-016-119364 (1-601)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 218

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 158

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 98

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          97 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 38

Qy          81 LeuThrProLeuGlnThrHisLeuThrMetLysGly 92
               ||||||||||||||||||||||||||||||||||
Db          37 CTCACACCCTTGCAGACACACTTGACCATGAAAGGT 2

RESULT 8
US-09-248-796A-148
; Sequence 148, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 148
;   LENGTH: 2472
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-148

Alignment Scores:
Pred. No.:                0.697          Length:         2472
Score:                    83.50          Matches:        35
Percent Similarity:       44.17%         Conservative:   18
Best Local Similarity:    29.17%         Mismatches:     52
Query Match:              12.03%         Indels:         15
DB:                       4              Gaps:           5

US-09-974-546C-84 (1-135) x US-09-248-796A-148 (1-2472)

Qy          24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
               |||::::::::::   |||   |||   |||   |||       :::|||:::       |||
Db        1634 AAAATCCAAAATCTACTACTGAATCTTCCATGGCTATCTTGTCTGCTGTCGGTGGTGACG 1693

Qy          44 ValLeuLeuProPheCys-----------TyrLysValPheArgLysLysGluLysVal 59
               :::::::::                 ||||||   :::||||||   |||
Db        1694 ATGATGATGACGATGATGAAGAAGAAGAAGATGAAGGTGATGAAGAAGAAGAGGAAG--- 1750

Qy          60 LysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAla 79
               |||      :::|||      :::|||   |||   :::||||||   |||
Db        1751 AAGATGAAGAAGAAGACGATCCAGTTTGGAGATTCAACGATAGAGATGTCATTTTGTACA 1810

Qy          80 Ile------LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCys 97
               |||      |||   ||||||      :::   ||||||          :::
Db        1811 ATATTGCTCTTGGTGCCACTACCAAACAATTGCACTATGTTTATGAAAATGATTCTGATT 1870

Qy          98 SerSerLeuSer-----------------SerGluAlaIleLeuPheThrLeuThrLeu 111
               |||   |||                  |||   :::||||||      |||
Db        1871 TCCAAGTTATTCCAACTTTTGGTCATTTGATCACCTTCAATTCTGGTAAATCTCAAAACT 1930

Qy         112 GlnLeuThrGln------ThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysThr 129
                  |||                  :::||||||         ::::::   |||
Db        1931 CCTTTGCTAAATTGTTACGTAACTTCAACCCAATGTTGTTATTACATGGTGAACATTACT 1990

RESULT 9
US-09-186-276B-53/c
; Sequence 53, Application US/09186276B
; Patent No. 6388173
; GENERAL INFORMATION:
;  APPLICANT: Benfey, Philip
;  APPLICANT:  DiLaurenzio, Laura
;  APPLICANT:  Wysocka-Diller, Joanna
;  APPLICANT:  Malamy, Jocelyn E.
;  APPLICANT:  Pysh, Leonard
;  APPLICANT:  Helariutta, Yrjo
;  TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
;  FILE REFERENCE: 5914-075-999
;  CURRENT APPLICATION NUMBER: US/09/186,276B
;  CURRENT FILING DATE:  1998-11-05
;  PRIOR APPLICATION NUMBER: 08/842,445
;  PRIOR FILING DATE: 1997-04-24
;  PRIOR APPLICATION NUMBER: 08/638,617
;  PRIOR FILING DATE: 1996-04-26
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
;   LENGTH: 1093
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-186-276B-53

Alignment Scores:
Pred. No.:                0.5            Length:         1093
Score:                    80.50          Matches:        37
Percent Similarity:       36.81%         Conservative:   16
Best Local Similarity:    25.69%         Mismatches:     47
Query Match:              11.60%         Indels:         44
DB:                       3              Gaps:           6

US-09-974-546C-84 (1-135) x US-09-186-276B-53 (1-1093)

Qy          10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
               |||   |||             :::   :::      |||             |||
Db         921 TACAGCGACAGATGTGTTTTTGTTGTGCTCTTCTTTATCCACTCAAGTAACCATCATCAC 862

Qy          30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
                  |||   |||   :::      :::      |||                  |||
Db         861 TATAACTTACCCAATATAGTTTGGGTTTGGCTTCCACACGGAACA-----------TGT 814

Qy          50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
                                    |||:::|||:::||||||      |||::::::|||:::
Db         813 AGCCATGGGTCGTCTCTTCCAGAAGAGGTAGAGCGCTCCTTCATGGCCTCCCAGTTTGTA 754

Qy          70 AspTyrSerIle---------------------------------------------- 73
               :::   |||
Db         753 GTTTTTGTCATAAGCTTTCAGCATCTCACTCGCTGCAAACGCTGCAGATGTGCTGACCGG 694

Qy          74 ---------GluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHis-LeuThrMe 90
                        |||   |||||||||                          ||| ||||||::
Db         693 CCAACCCGTGAACCCAGCCATCAT----------------------CATTCTGACCCT 658

Qy          90 tLysGlySerSerMetLysCysSerSerLeu---------SerSerGluAlaIleLeuPh 107
               :          |||      |||   |||            ||||||   ||||||:::||
Db         657 CCATTTCCCCAGTACCTCGTGTCTCTCTACTCTCTCTGACTCCTCACATGCTATCATGTT 598

Qy         107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys---------LeuLe 124
               ||||:::                 :::|||      :::||||||         :::||
Db         597 CACTATG-----------------TCTCTTGCTACACAGTGTTGTTCTGCGCTGATTCT 556

Qy         124 uTyrLeuSer 127
               |   ||||||
Db         555 CTGCTTATCA 546

RESULT 10
US-08-842-445-53/c
; Sequence 53, Application US/08842445A
; Patent No. 6441270
; GENERAL INFORMATION:
;  APPLICANT: Benfey et al.
;  TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 5914-056-999
;  CURRENT APPLICATION NUMBER: US/08/842,445A
;  CURRENT FILING DATE: 1997-04-24
;  EARLIER APPLICATION NUMBER: 08/638,617
;  EARLIER FILING DATE: 1996-04-26
;  NUMBER OF SEQ ID NOS: 79
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
;   LENGTH: 1093
;   TYPE: DNA
;   ORGANISM: Plant
US-08-842-445-53

Alignment Scores:
Pred. No.:              0.5          Length:        1093
Score:                  80.50        Matches:       37
Percent Similarity:     36.81%       Conservative:  16
Best Local Similarity:  25.69%       Mismatches:    47
Query Match:            11.60%       Indels:        44
DB:                     3            Gaps:          6

US-09-974-546C-84 (1-135) x US-08-842-445-53 (1-1093)

Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              |||   |||         :::   :::      |||              |||
Db        921 TACAGCGACAGATGTGTTTTTGTTGTGCTCTTCTTTATCCACTCAAGTAACCATCATCAC 862

Qy         30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
                 |||   |||   :::      :::      |||                    |||
Db        861 TATAACTTACCCAATATAGTTTGGGTTTGGCTTCCACACGGAACA-----------TGT 814

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
                           |||:::|||:::||||||      |||::::::|||:::
Db        813 AGCCATGGGTCGTCTCTTCCAGAAGAGGTAGAGCGCTCCTTCATGGCCTCCCAGTTTGTA 754

Qy         70 AspTyrSerIle---------------------------------------------- 73
              :::   |||
Db        753 GTTTTTGTCATAAGCTTTCAGCATCTCACTCGCTGCAAACGCTGCAGATGTGCTGACCGG 694

Qy         74 ---------GluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHis-LeuThrMe 90
                      |||   |||||||||                          ||| ||||||::
Db        693 CCAACCCGTGAACCCAGCCATCAT----------------------CATTCTGACCCT 658

Qy         90 tLysGlySerSerMetLysCysSerSerLeu---------SerSerGluAlaIleLeuPh 107
              :         |||      |||   |||            ||||||   ||||||:::||
Db        657 CCATTTCCCCAGTACCTCGTGTCTCTCTACTCTCTCTGACTCCTCACATGCTATCATGTT 598

Qy        107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys---------LeuLe 124
              ||||:::                :::|||     :::||||||         :::||
Db        597 CACTATG-----------------TCTCTTGCTACACAGTGTTGTTCTGCGCTGATTCT 556

Qy        124 uTyrLeuSer 127
              |   ||||||
Db        555 CTGCTTATCA 546

RESULT 11
US-09-186-188B-53/c
; Sequence 53, Application US/09186188B
; Patent No. 6455672
; GENERAL INFORMATION:
;  APPLICANT: Benfey et al.
;  TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE: 5914-074-999
;  CURRENT APPLICATION NUMBER: US/09/186,188B
;  CURRENT FILING DATE:  1998-11-05
;  PRIOR APPLICATION NUMBER: 08/842,445
;  PRIOR FILING DATE: 1997-04-24
;  PRIOR APPLICATION NUMBER: 08/638,617
;  PRIOR FILING DATE: 1996-04-26
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
;   LENGTH: 1093
;   TYPE: DNA
;   ORGANISM: Plant
US-09-186-188B-53

Alignment Scores:
```

```
Pred. No.:              0.5          Length:        1093
Score:                  80.50        Matches:       37
Percent Similarity:     36.81%       Conservative:  16
Best Local Similarity:  25.69%       Mismatches:    47
Query Match:            11.60%       Indels:        44
DB:                     3            Gaps:          6

US-09-974-546C-84 (1-135) x US-09-186-188B-53 (1-1093)

Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              |||   |||         :::   :::      |||              |||
Db        921 TACAGCGACAGATGTGTTTTTGTTGTGCTCTTCTTTATCCACTCAAGTAACCATCATCAC 862

Qy         30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
                 |||   |||   :::      :::      |||                    |||
Db        861 TATAACTTACCCAATATAGTTTGGGTTTGGCTTCCACACGGAACA-----------TGT 814

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
                           |||:::|||:::||||||      |||::::::|||:::
Db        813 AGCCATGGGTCGTCTCTTCCAGAAGAGGTAGAGCGCTCCTTCATGGCCTCCCAGTTTGTA 754

Qy         70 AspTyrSerIle---------------------------------------------- 73
              :::   |||
Db        753 GTTTTTGTCATAAGCTTTCAGCATCTCACTCGCTGCAAACGCTGCAGATGTGCTGACCGG 694

Qy         74 ---------GluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHis-LeuThrMe 90
                      |||   |||||||||                          ||| ||||||::
Db        693 CCAACCCGTGAACCCAGCCATCAT----------------------CATTCTGACCCT 658

Qy         90 tLysGlySerSerMetLysCysSerSerLeu---------SerSerGluAlaIleLeuPh 107
              :         |||      |||   |||            ||||||   ||||||:::||
Db        657 CCATTTCCCCAGTACCTCGTGTCTCTCTACTCTCTCTGACTCCTCACATGCTATCATGTT 598

Qy        107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys---------LeuLe 124
              ||||:::                :::|||     :::||||||         :::||
Db        597 CACTATG-----------------TCTCTTGCTACACAGTGTTGTTCTGCGCTGATTCT 556

Qy        124 uTyrLeuSer 127
              |   ||||||
Db        555 CTGCTTATCA 546

RESULT 12
US-09-265-585C-53/c
; Sequence 53, Application US/09265585C
; Patent No. 6809234
; GENERAL INFORMATION:
;  APPLICANT: Benfey, Phillip N.
;  APPLICANT:  Di Laurenzio, Laura
;  APPLICANT:  Wysocka-Diller, Joanna
;  APPLICANT:  Malamy, Jocelyn E.
;  APPLICANT:  Pysh, Leonard
;  APPLICANT:  Helariutta, Yrjo
;  APPLICANT:  Bruce, Wesley
;  APPLICANT:  Lim, Jun
;  TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
;  FILE REFERENCE: 5914-066
;  CURRENT APPLICATION NUMBER: US/09/265,585C
;  CURRENT FILING DATE:  1999-03-10
;  PRIOR APPLICATION NUMBER: 08/842,445
;  PRIOR FILING DATE: 1997-04-24
;  PRIOR APPLICATION NUMBER: 08/638,617
;  PRIOR FILING DATE: 1996-04-26
;  NUMBER OF SEQ ID NOS: 152
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 1093
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-265-585C-53

Alignment Scores:
Pred. No.:              0.5          Length:        1093
Score:                  80.50        Matches:       37
```

```
Percent Similarity:     36.81%          Conservative:  16
Best Local Similarity:  25.69%          Mismatches:    47
Query Match:            11.60%          Indels:        44
DB:                     4               Gaps:          6

US-09-974-546C-84 (1-135) x US-09-265-585C-53 (1-1093)

Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              |||   |||          :::   :::      |||           |||
Db        921 TACAGCGACAGATGTGTTTTTGTTGTGCTCTTCTTTATCCACTCAAGTAACCATCATCAC 862

Qy         30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
                 |||   |||   :::      :::      |||                     |||
Db        861 TATAACTTACCCAATATAGTTTGGGTTTGGCTTCCACACGGAACA-----------TGT 814

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
                           |||:::|||:::||||||      |||::::::|||:::
Db        813 AGCCATGGGTCGTCTCTTCCAGAAGAGGTAGAGCGCTCCTTCATGGCCTCCCAGTTTGTA 754

Qy         70 AspTyrSerIle---------------------------------------------- 73
                 :::   |||
Db        753 GTTTTTGTCATAAGCTTTCAGCATCTCACTCGCTGCAAACGCTGCAGATGTGCTGACCGG 694

Qy         74 ---------GluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHis-LeuThrMe 90
                      |||   |||||||||                        ||| ||||||::
Db        693 CCAACCCGTGAACCCAGCCATCAT-----------------------CATTCTGACCCT 658

Qy         90 tLysGlySerSerMetLysCysSerSerLeu---------SerSerGluAlaIleLeuPh 107
              :       |||     |||   |||              ||||||  ||||||:::||
Db        657 CCATTTCCCCAGTACCTCGTGTCTCTCTACTCTCTCTGACTCCTCACATGCTATCATGTT 598

Qy        107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys---------LeuLe 124
              ||||:::                  :::|||      :::|||||          :::||
Db        597 CACTATG-----------------TCTCTTGCTACACAGTGTTGTTCTGCGCTGATTCT 556

Qy        124 uTyrLeuSer 127
              |   ||||||
Db        555 CTGCTTATCA 546

RESULT 13
US-09-949-016-15729/c
; Sequence 15729, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15729
;   LENGTH: 148783
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(148783)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15729

Alignment Scores:
Pred. No.:              1.37e+03        Length:        148783
Score:                  80.00           Matches:       41
Percent Similarity:     41.48%          Conservative:  15
Best Local Similarity:  30.37%          Mismatches:    52

Query Match:            11.53%          Indels:        29
DB:                     4               Gaps:          4

US-09-974-546C-84 (1-135) x US-09-949-016-15729 (1-148783)

Qy          5 LeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLys 24
              |||   :::      |||:::|||          |||||||||||||||||||||||||
Db      15651 CTTCATGACAATCCATATAAAGATACCACTATTATTATCCACATTTTACAGATAAGAAAA 15592

Qy         25 LeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThrVal 44
              ||||||||||||      |||      :::   |||   :::::::::
Db      15591 CTGAGGCACAGAGA-CTAAATAAATCTAAATTCCCACAGCTAGTAAGT------------ 15545

Qy         45 LeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGlnLys 64
                                                    :::   |||         :::
Db      15544 ----------------------------------GGCAGAGCCAAGAATTAAGCTCAG 15521

Qy         65 AlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrProLeu 84
              ||||||   |||          |||              |||:::   :::         |||
Db      15520 GCCACCTGATTC------TCCAGCTTCTACCCAGTACACTATACCGTGTTCCCATG-CTT 15468

Qy         85 GlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSer--------- 101
                 ||||||      |||:::          :::      |||:::
Db      15467 TCAACGCATTCCCAAATGGAGCAGGGGAGAATAAGCGCCAGCAACCAACAAAGGAGACAT 15408

Qy        102 ------------SerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeu 117
                         |||             ||||||:::|||      |||      |||
Db      15407 CGACGCCAAAGGTCTACCATAAAGTTTTTCACGGTGACATCCTTGCTT------ACCTGT 15354

Qy        118 GlyLeuGluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
                 |||   |||   |||   |||:::|||       |||
Db      15353 CCAGGGCCATGCTGGTTGGCATACTCCTCCCAAACCCGCGGACCC 15309

RESULT 14
US-09-573-080A-99/c
; Sequence 99, Application US/09573080A
; Patent No. 6828097
; GENERAL INFORMATION:
;  APPLICANT:  JOAN, KNOLL
;  APPLICANT:  ROGAN, PETER
;  TITLE OF INVENTION:   SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATIN
;  FILE REFERENCE:   30307
;  CURRENT APPLICATION NUMBER: US/09/573,080A
;  CURRENT FILING DATE:  2000-05-16
;  NUMBER OF SEQ ID NOS:   479
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 99
;   LENGTH: 262
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(262)
;   OTHER INFORMATION: mir
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;   AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;   TITLE: Prototypic sequences for human repetitive DNA
;   JOURNAL: Journal of Molecular Evolution
;   VOLUME: 35
;   ISSUE: 4
;   PAGES: 286-291
;   DATE: 1992-10-__
;   DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;   DATABASE ENTRY DATE:  ____-__-__
;   DATABASE ENTRY DATE: 1996-01-26
US-09-573-080A-99

Alignment Scores:
Pred. No.:              0.0721          Length:        262
Score:                  79.50           Matches:       27
Percent Similarity:     46.67%          Conservative:  1
Best Local Similarity:  45.00%          Mismatches:    10
```

```
Query Match:            11.46%          Indels:         22
DB:                     4               Gaps:           2

US-09-974-546C-84 (1-135) x US-09-573-080A-99 (1-262)

Qy          8 GlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHis 27
              |||   ||||||             ||||||   |||||||||:::|||||||||||||||
Db        174 CAACCCTATGAGGTAGGTACTATTATTATCCCCATTTTACAGATGAGGAAACTGAGGCAC 115

Qy         28 ArgLeu-------------------------------------------------------Ser 30
              |||                                                          |||
Db        114 AGAGAGGTTAAGTAACTTGCCCAAGGTCACACAGCTAGTAAGTGGCAGAGCCGGGATTCG 55

Qy         31 AsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeu-ProPheCys 49
              |||            ||||||   ||||||||||||   ||||||||| |||   |||
Db         54 AAC-----------CCAGGCAGTCTGGCTCCAGAGYCCGTGCTCTTAACCACTATGC 9

RESULT 15
US-08-675-566-5
; Sequence 5, Application US/08675566
; Patent No. 6090393
;  GENERAL INFORMATION:
;    APPLICANT:  Fischer, Laurent
;    TITLE OF INVENTION:  PROMOTERS, EXPRESSION CASSETTES,
;    TITLE OF INVENTION:  RECOMBINANT VIRUSES, METHODS FOR MAKING, AND USES THEREOF
;    NUMBER OF SEQUENCES:  120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/675,566
;      FILING DATE:  03-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer Esq., William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2890
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)840-3333
;      TELEFAX:  (212)840-0712
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6196 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-675-566-5

Alignment Scores:
Pred. No.:              10.6            Length:         6196
Score:                  79.50           Matches:        30
Percent Similarity:     46.40%          Conservative:   28
Best Local Similarity:  24.00%          Mismatches:     48
Query Match:            11.46%          Indels:         19
DB:                     3               Gaps:           4

US-09-974-546C-84 (1-135) x US-08-675-566-5 (1-6196)

Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              |||:::|||:::|||   ::::::               |||:::|||   |||
Db       1747 TACAAAGACCTGCATCCCCTCCTTAGC-----------AGACAGCTCTTGCACACACGC 1794

Qy         30 SerAsnPhePro-ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCy 49
              ||||||:::||| ||||||         |||:::|||      :::      |||
Db       1795 AGTAACTATCCACCGCTTAAGAAAAGCTTTAAGCCCAGCGCACATAACAGCTCCAATG-- 1852

Qy         49 sTyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIl 69
                          |||:::   |||:::      |||      :::|||      ::
Db       1853 -------TTTTTATCCAAGGAGAGCAAAATTTCAGCAAGCGCAGGCTCAACAGTAATAGT 1905

Qy         69 e-AspTyrSerIleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuT 89
              :          |||            |||   :::   :::|||
Db       1906 GAAGCAGAGGCATTTCAGACGAGGCTCACTAGCTGCAGTCGCCATTTATGAGG------- 1958

Qy         89 hrMetLysGlySerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrL 109
                        |||:::::::|||   :::|||   ::::::|||   :::   |||
Db       1959 ----------TCTGCAATAAAAAACAACTCATCAGCAGCTGAAAAAGTGCACTTT---- 2003

Qy        109 euThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysT 129
                        |||   :::   |||   ::::::      :::         ::::::|
Db       2004 --------GACCTCATTAAGCCACTGCATATGCAAGTCCTCATCTATGCCGCAGCCCAGA 2055

Qy        129 hrIleHisPro 132
              ||::::::|||
Db       2056 CCCTCAATCCA 2066


Search completed: September  9, 2005, 17:01:29
Job time : 139.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 14:46:00 ; Search time 408.5 Seconds
(without alignments)
2170.324 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174504_21077/app_query.fasta_1.654
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09974546_@CGN_1_1_456_@runat_07092005_174504_21077
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		694	100.0	2087	10	US-09-974-546-83	Sequence 83, Appl
2		694	100.0	2505	10	US-09-974-546-85	Sequence 85, Appl
3		90	13.0	1113	17	US-10-369-493-44085	Sequence 44085, A
4		90	13.0	1146	17	US-10-369-493-25021	Sequence 25021, A
5		87	12.5	1258	18	US-10-425-114-318	Sequence 318, App
6		87	12.5	1624	20	US-10-425-115-49918	Sequence 49918, A
7	c	83	12.0	146547	15	US-10-017-128-1	Sequence 1, Appli
8	c	82	11.8	24601	9	US-09-070-927A-223	Sequence 223, App
9	c	81.5	11.7	133462	19	US-10-367-094-114	Sequence 114, App
10		81	11.7	381	9	US-09-960-352-9738	Sequence 9738, Ap
11		81	11.7	819	19	US-10-437-963-23499	Sequence 23499, A
12	c	81	11.7	1277	20	US-10-425-115-137953	Sequence 137953,
13		81	11.7	4437	20	US-10-425-115-178334	Sequence 178334,
14	c	80.5	11.6	455	20	US-10-425-115-63753	Sequence 63753, A
15	c	80.5	11.6	600	22	US-10-972-079-4637	Sequence 4637, Ap
16	c	80.5	11.6	600	22	US-10-972-079-4638	Sequence 4638, Ap
17	c	80.5	11.6	1026	18	US-10-425-114-13022	Sequence 13022, A
18	c	80.5	11.6	1093	9	US-09-186-276B-53	Sequence 53, Appl
19	c	80.5	11.6	1093	9	US-09-186-188B-53	Sequence 53, Appl
20	c	80.5	11.6	1093	14	US-10-253-007-53	Sequence 53, Appl
21		80.5	11.6	2343	20	US-10-425-115-127057	Sequence 127057,
22		80	11.5	681	19	US-10-437-963-98199	Sequence 98199, A
23		80	11.5	1355	19	US-10-437-963-37754	Sequence 37754, A
24		80	11.5	1693	20	US-10-723-860-7296	Sequence 7296, Ap
25	c	79.5	11.5	262	10	US-09-854-867-99	Sequence 99, Appl
26	c	79.5	11.5	262	21	US-10-786-970A-99	Sequence 99, Appl
27		79.5	11.5	567	18	US-10-424-599-74400	Sequence 74400, A
28		79.5	11.5	585	20	US-10-357-930-54224	Sequence 54224, A
29		79	11.4	590	13	US-10-027-632-269701	Sequence 269701,
30		79	11.4	590	17	US-10-027-632-269701	Sequence 269701,
31		79	11.4	831	13	US-10-027-632-133030	Sequence 133030,
32		79	11.4	831	17	US-10-027-632-133030	Sequence 133030,
33	c	79	11.4	4068	19	US-10-437-963-35166	Sequence 35166, A
34	c	79	11.4	4107	20	US-10-425-115-139445	Sequence 139445,
35		79	11.4	4335	20	US-10-425-115-139432	Sequence 139432,
36	c	78.5	11.3	386	9	US-09-960-352-10443	Sequence 10443, A
37		78.5	11.3	600	22	US-10-972-079-88625	Sequence 88625, A
38	c	78.5	11.3	1787	13	US-10-027-632-84669	Sequence 84669, A
39	c	78.5	11.3	1787	13	US-10-027-632-84670	Sequence 84670, A
40	c	78.5	11.3	1787	17	US-10-027-632-84669	Sequence 84669, A
41	c	78.5	11.3	1787	17	US-10-027-632-84670	Sequence 84670, A
42	c	78	11.2	650	13	US-10-027-632-223621	Sequence 223621,
43	c	78	11.2	650	17	US-10-027-632-223621	Sequence 223621,
44		78	11.2	1207	18	US-10-424-599-5755	Sequence 5755, Ap
45	c	78	11.2	9869	20	US-10-425-115-139430	Sequence 139430,

ALIGNMENTS

RESULT 1
US-09-974-546-83
; Sequence 83, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```
;               STATE: Texas
;               COUNTRY: USA
;               ZIP: 77210
;          COMPUTER READABLE FORM:
;               MEDIUM TYPE: Floppy disk
;               COMPUTER: IBM PC compatible
;               OPERATING SYSTEM: PC-DOS/MS-DOS
;               SOFTWARE: PatentIn Release #1.0, Version #1.30
;          CURRENT APPLICATION DATA:
;               APPLICATION NUMBER: US/09/974,546
;               FILING DATE: 10-Oct-2001
;               CLASSIFICATION: Unknown
;          PRIOR APPLICATION DATA:
;               APPLICATION NUMBER: 09/097,199
;               FILING DATE: 1998-06-12
;          ATTORNEY/AGENT INFORMATION:
;               NAME: Nakashima, Richard A.
;               REGISTRATION NUMBER: P-42,023
;               REFERENCE/DOCKET NUMBER: UROC:018
;          TELECOMMUNICATION INFORMATION:
;               TELEPHONE: (512) 418-3000
;               TELEFAX: (512) 474-7577
;     INFORMATION FOR SEQ ID NO: 83:
;          SEQUENCE CHARACTERISTICS:
;               LENGTH: 2087 base pairs
;               TYPE: nucleic acid
;               STRANDEDNESS: single
;               TOPOLOGY: linear
;          FEATURE:
;               NAME/KEY:  CDS
;               LOCATION:  99..503
;          SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-974-546-83

Alignment Scores:
Pred. No.:                5.82e-85        Length:         2087
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-84 (1-135) x US-09-974-546-83 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnProProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
```

```
US-09-974-546-85
; Sequence 85, Application US/09974546
; Publication No. US20030050470A1
;  GENERAL INFORMATION:
;       APPLICANT: An, Gang
;                  O'Hara, S. Mark
;                  Ralph, David
;                  Veltri, Robert
;       TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                           PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Arnold, White & Durkee
;            STREET: P.O. Box 4433
;            CITY: Houston
;            STATE: Texas
;            COUNTRY: USA
;            ZIP: 77210
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/974,546
;            FILING DATE: 10-Oct-2001
;            CLASSIFICATION: Unknown
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/097,199
;            FILING DATE: 1998-06-12
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Nakashima, Richard A.
;            REGISTRATION NUMBER: P-42,023
;            REFERENCE/DOCKET NUMBER: UROC:018
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (512) 418-3000
;            TELEFAX: (512) 474-7577
;  INFORMATION FOR SEQ ID NO: 85:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2505 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  99..503
;       SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-974-546-85

Alignment Scores:
Pred. No.:                7.74e-85        Length:         2505
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-84 (1-135) x US-09-974-546-85 (1-2505)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnProProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:09:35 ; Search time 1867.5 Seconds
(without alignments)
2751.634 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174502_20989/app_query.fasta_1.654
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_3556_@runat_07092005_174502_20989 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	97	14.0	690	5	BU242169	BU242169 603779592
	2	95.5	13.8	588	5	BW569427	BW569427 BW569427
	3	94	13.5	606	6	CB286029	CB286029 CMD29_B11
	4	90.5	13.0	459	5	BW557636	BW557636 BW557636
	5	90.5	13.0	543	5	BW582739	BW582739 BW582739
	6	90	13.0	570	5	BW591794	BW591794 BW591794
c	7	89.5	12.9	1151	5	BU413225	BU413225 603155489
c	8	89	12.8	811	5	BU436282	BU436282 603209819
	9	88.5	12.8	743	7	CV511381	CV511381 kc40h08.y
	10	88.5	12.8	789	8	BH559290	BH559290 BOHOW80TR
	11	88	12.7	776	9	AG377362	AG377362 Mus muscu
c	12	87.5	12.6	803	5	BX756993	BX756993 BX756993
c	13	87	12.5	788	9	CR262699	CR262699 Reverse s
	14	86.5	12.5	741	3	AY439721	AY439721 Armigeres
c	15	86	12.4	529	5	BQ592283	BQ592283 E012698-0
	16	86	12.4	769	8	BZ521554	BZ521554 BOKAE11TF
c	17	86	12.4	992	9	CL254017	CL254017 ZMMBBb060
c	18	85.5	12.3	788	8	CC091268	CC091268 CSU-K33r.
c	19	85.5	12.3	877	6	CA472302	CA472302 AGENCOURT
c	20	85	12.2	579	9	CE464906	CE464906 tigr-gss-
c	21	85	12.2	707	9	CC743728	CC743728 ZMMBBb011
	22	85	12.2	906	9	CG300818	CG300818 OG0AB25TV
c	23	85	12.2	914	9	CG300807	CG300807 OG0AB25TH
	24	84.5	12.2	285	4	BM111157	BM111157 EST558693
	25	84.5	12.2	562	8	AZ954412	AZ954412 2M0220O05
c	26	84.5	12.2	652	2	BF644675	BF644675 NF018F01E
c	27	84.5	12.2	688	5	BP027905	BP027905 BP027905
c	28	84.5	12.2	741	4	BI309709	BI309709 EST531119
c	29	84.5	12.2	770	4	BM780176	BM780176 EST590764
	30	84	12.1	493	5	BX836326	BX836326 BX836326
	31	84	12.1	624	8	BZ745987	BZ745987 OGFBK05TM
c	32	84	12.1	728	6	CD821188	CD821188 BN25.041B
	33	84	12.1	896	9	CG008627	CG008627 ZUACO85TH
	34	84	12.1	916	7	CK151729	CK151729 FGAS03441
c	35	84	12.1	969	2	BE739923	BE739923 601556175
c	36	84	12.1	1399	2	AW730867	AW730867 GA__Ea002
	37	83.5	12.0	721	8	B69065	B69065 CIT-HSP-205
	38	83.5	12.0	732	5	BU264381	BU264381 603505585
	39	83.5	12.0	781	7	CN135272	CN135272 OX1_31_E0
	40	83.5	12.0	952	1	AL665968	AL665968 AL665968
	41	83.5	12.0	1335	4	BM467342	BM467342 AGENCOURT
	42	83	12.0	395	6	CB069753	CB069753 is28d02.x
	43	83	12.0	572	2	BE015616	BE015616 EST842 Ma
	44	83	12.0	612	2	BE015598	BE015598 EST824 Ma
	45	83	12.0	634	9	AG110445	AG110445 Pan trogl

ALIGNMENTS

RESULT 1
BU242169/c
LOCUS BU242169 690 bp mRNA linear EST 26-NOV-2002
DEFINITION 603779592F1 CSEQCHN34 Gallus gallus cDNA clone ChEST725h23 5', mRNA sequence.
ACCESSION BU242169
VERSION BU242169.1 GI:25488447
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 690)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES Location/Qualifiers
source 1. .690
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 15I"

/db_xref="taxon:9031"
/clone="ChEST725h23"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN34"
/note="Organ: liver; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: Not1; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Alignment Scores:

Pred. No.:	1.02	Length:	690
Score:	97.00	Matches:	33
Percent Similarity:	45.08%	Conservative:	22
Best Local Similarity:	27.05%	Mismatches:	59
Query Match:	13.98%	Indels:	8
DB:	5	Gaps:	3

US-09-974-546C-84 (1-135) x BU242169 (1-690)

```
Qy         15 AsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArg 34
              |||:::      ||||||:::         :::         :::||||||
Db        623 AACTTAAATAGCATTTTACGTTTCCCTACAGTTACTAAAGAAATATCAAAT------GAT 570

Qy         35 LeuProGlyIleLeuAlaPro---------GluThrValLeuLeuProPheCysTyrLys 51
              :::||||||:::|||   |||                    ||||||
Db        569 GTCCCAGGACTCCTGTTGCCCACAGAGGAGAACAGGTATTTGCTCAGCACTATCACAGTG 510

Qy         52 ValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyr 71
                ||||||   |||   :::         :::   ||||||   |||
Db        509 AAATTTAGGTGAAAATAAAGAACAGTATACACTGATAAAGCTGCAGAG---------AAT 459

Qy         72 SerIleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLys 91
              |||:::   :::   |||   :::      :::      :::      |||   ::::::
Db        458 TCAGTTATAAAAGGCCATTTCTCATTTCAGAGCTTTGCTAGGATGCCTCTTCTTATAAGA 399

Qy         92 GlySerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                    :::   :::         :::|||   :::||||||
Db        398 AAAAAAGGAAAAAGAAAAGCACCACCACAAACTGAGATGCTGCTATTTCTCACTAAATGC 339

Qy        112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysThrIleHis 131
              :::   |||   |||           ||||||||||||:::         |||||||||
Db        338 AAGGCCACCACTACATCCTCTTGTTTCTGTTGCCTTCTGTTTACAGGAGTAACCATTCAT 279

Qy        132 ProGln 133
              |||:::
Db        278 CCAAAG 273
```

RESULT 2
BW569427

LOCUS	BW569427	588 bp	mRNA	linear	EST 31-AUG-2004

DEFINITION BW569427 Yutaka Satou unpublished cDNA library (cstb) Ciona savignyi cDNA clone cstb014a08 3', mRNA sequence.
ACCESSION BW569427
VERSION BW569427.1 GI:51749861
KEYWORDS EST.
SOURCE Ciona savignyi
ORGANISM Ciona savignyi
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 588)
AUTHORS Satou,Y. and Satoh,N.
TITLE Expressed genes in Ciona savignyi
JOURNAL Unpublished (2004)
COMMENT Contact: Yutaka Satou
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: yutaka@ascidian.zool.kyoto-u.ac.jp.
FEATURES Location/Qualifiers
source 1. .588
/organism="Ciona savignyi"
/mol_type="mRNA"
/db_xref="taxon:51511"
/clone="cstb014a08"
/dev_stage="tailbud stage"
/clone_lib="Yutaka Satou unpublished cDNA library (cstb)"
ORIGIN

Alignment Scores:

Pred. No.:	1.23	Length:	588
Score:	95.50	Matches:	37
Percent Similarity:	44.27%	Conservative:	21
Best Local Similarity:	28.24%	Mismatches:	48
Query Match:	13.76%	Indels:	25
DB:	5	Gaps:	6

US-09-974-546C-84 (1-135) x BW569427 (1-588)

```
Qy         24 LysLeuArgHisArgLeuSerAsn----------ProProArgLeuProGlyIleLeu 39
              |||       |||:::|||   |||            |||   :::|||      :::
Db         82 AAATATTCACATAAACTTGGCAACAAAAACACAAGATTTTTACAACTTGCTCTAGTTTGA 141

Qy         40 AlaProGluThrValLeuLeuProPheCysTyrLysValPhe----------ArgLys 55
                 |||::::::   |||          |||::::::
Db        142 CTAGAAGAATCCATTTTTTTAAAGGACGAAAGCAAAATTTACATAATTTGCAATAATCCG 201

Qy         56 LysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGln 75
              |||   ::::::|||   ::::::   ::::::|||::::::||||||   ||||||
Db        202 TAGGAATTGCTTAGAAGATTTAAAAGAAGAAGCGATTTCGTAGAATATAGCACGGAACAA 261

Qy         76 Ser------HisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySer 93
              |||       ||||||||||||         :::|||:::   |||   :::   ||||||
Db        262 AGCAATTTGCACCACGCTATATTTGCAGGGATTCAGTCATTTCTTCCACTGAGTGGCTCA 321

Qy         94 SerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeu 113
                    |||         |||         |||                  ::::::
Db        322 AGAGAGCACTGTTGCTTGTGCAGCTGCTGATACTGCCTC----------CAGCGGATT 369

Qy        114 ThrGlnThrLeuGlyLeuGluCys--------------------CysLeuLeu------ 124
                   :::||||||   |||                      |||   |||
Db        370 GGGACTGGGGTGGGTCTGTTGTGCATTCGTCAAGAATTGTGAGGGTGCTCGTTGGTCCAC 429

Qy        125 ------TyrLeuSerLysThrIleHisProGln 133
                   ||||||               ||||||
Db        430 TACTGCTACCTTGTGGATGCTTTTCCTCCACAA 462
```

RESULT 3
CB286029

LOCUS	CB286029	606 bp	mRNA	linear	EST 27-FEB-2003

DEFINITION CMD29_B11_90 UMNMPM3 Sus scrofa cDNA clone PPSUBLIB_15B11 5', mRNA sequence.
ACCESSION CB286029
VERSION CB286029.1 GI:28576983
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 606)
AUTHORS Dvorak,C.M.T., Hyland,K.A., Zhang,Y., Fahrenkrug,S.C. and Murtaugh,M.P.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:23:48 ; Search time 58 Seconds
(without alignments)
900.218 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	3	AAY59296	Aay59296 Prostate
2	694	100.0	135	3	AAY59295	Aay59295 Prostate
3	694	100.0	135	4	AAU02174	Aau02174 Biomarker
4	694	100.0	135	4	AAU02175	Aau02175 Biomarker
5	103	14.8	21	3	AAY59294	Aay59294 Cancer bi
6	103	14.8	21	4	AAU02173	Aau02173 Biomarker
7	90	13.0	371	8	ADS41968	Ads41968 Bacterial
8	90	13.0	381	8	ADN18681	Adn18681 Bacterial
9	82.5	11.9	645	5	ABB90843	Abb90843 Herbicida
10	81	11.7	384	8	ADN48099	Adn48099 Thermococ
11	77.5	11.2	1061	7	ADF50279	Adf50279 Human PFM
12	77.5	11.2	1061	8	ADO36741	Ado36741 Human PR-
13	77.5	11.2	1144	4	ABB58124	Abb58124 Drosophil
14	77	11.1	1026	8	ADS08184	Ads08184 Staphyloc
15	76.5	11.0	322	6	ABM69043	Abm69043 Photorhab
16	75	10.8	633	6	ABU42279	Abu42279 Protein e
17	73	10.5	195	4	ABG18410	Abg18410 Novel hum
18	72	10.4	523	4	AAB46398	Aab46398 H. pylori
19	72	10.4	636	8	ADM94292	Adm94292 Soybean E
20	71.5	10.3	450	7	ADF41696	Adf41696 Bacillus
21	71.5	10.3	634	8	ADM94288	Adm94288 Corn Ethy
22	71.5	10.3	676	4	ABG15498	Abg15498 Novel hum
23	71	10.2	127	4	AAU42328	Aau42328 Propionib
24	71	10.2	127	6	ABM38847	Abm38847 Propionib
25	71	10.2	424	4	AAM39676	Aam39676 Human pol
26	71	10.2	424	4	AAG81358	Aag81358 Human AFP
27	71	10.2	424	8	ADH09599	Adh09599 Human hos
28	71	10.2	424	8	ADH09598	Adh09598 Human hos
29	71	10.2	424	8	ADH09600	Adh09600 Human hos
30	71	10.2	429	4	AAM41462	Aam41462 Human pol
31	71	10.2	519	4	AAB92584	Aab92584 Human pro
32	71	10.2	1922	8	ADL97802	Adl97802 Human dcr
33	71	10.2	1924	4	AAE09768	Aae09768 Human dic
34	71	10.2	1924	5	AAO15987	Aao15987 Human hel
35	71	10.2	1924	6	ABO07204	Abo07204 Human p53
36	71	10.2	1924	7	ABU64479	Abu64479 Stem loop
37	71	10.2	1924	7	ADN60159	Adn60159 Human hel
38	70.5	10.2	438	7	ABO64633	Abo64633 Klebsiell
39	70.5	10.2	634	8	ADM94300	Adm94300 Corn Ethy
40	70	10.1	206	7	ADJ69892	Adj69892 Human hea
41	70	10.1	982	8	ADS29846	Ads29846 Bacterial
42	69.5	10.0	1294	2	AAW30601	Aaw30601 Human typ
43	69.5	10.0	1353	8	ADQ89904	Adq89904 Antagonis
44	69.5	10.0	1398	3	AAB18292	Aab18292 Plasmodiu
45	69.5	10.0	3105	5	ABB80604	Abb80604 Human sbg

ALIGNMENTS

RESULT 1
AAY59296
ID AAY59296 standard; peptide; 135 AA.
XX
AC AAY59296;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AAZ87584.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 184-186; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used

CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 135 AA;

Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135
```

RESULT 2
AAY59295
ID AAY59295 standard; peptide; 135 AA.
XX
AC AAY59295;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AAZ87583.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 182-183; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 135 AA;

Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135
```

RESULT 3
AAU02174
ID AAU02174 standard; protein; 135 AA.
XX
AC AAU02174;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker protein encoded by UC band 28 #2, used in diagnosis of cancer.
XX
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human.
XX
OS Homo sapiens.
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
XX
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri R;
XX
DR WPI; 2001-289849/30.
DR N-PSDB; AAS04000.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS Disclosure; Col 121; 78pp; English.
XX
CC The sequence represents the amino acid sequence of biomarker protein
CC encoded by UC band 28 #2. Proteins encoded by the nucleic acid markers
CC can be used to produce antibodies for the detection of prostate, breast
CC or bladder cancer. Biomarker nucleic acid sequences can be used as
CC hybridisation probes and primers that specifically hybridise to prostate
CC cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
CC cancer markers. The nucleic acids can be used as targets for therapeutic
CC intervention in these diseases, in the identification and isolation of
CC full-length gene sequences, including regulatory elements for gene
CC expression, from genomic human DNA libraries, as hybridisation probes for
CC screening genomic human DNA libraries. The kits comprising the nucleic
CC acid sequences are useful for detecting bladder, breast or prostate

```
CC   cancer cells in a biological sample
XX
SQ   Sequence 135 AA;

  Query Match            100.0%;  Score 694;  DB 4;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-75;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 4
AAU02175
ID   AAU02175 standard; protein; 135 AA.
XX
AC   AAU02175;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker protein encoded by UC band 28 #3, used in diagnosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   N-PSDB; AAS04001.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Disclosure; Col 125; 78pp; English.
XX
CC   The sequence represents the amino acid sequence of biomarker protein
CC   encoded by UC band 28 #3. Proteins encoded by the nucleic acid markers
CC   can be used to produce antibodies for the detection of prostate, breast
CC   or bladder cancer. Biomarker nucleic acid sequences can be used as
CC   hybridisation probes and primers that specifically hybridise to prostate
CC   cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
CC   cancer markers. The nucleic acids can be used as targets for therapeutic
CC   intervention in these diseases, in the identification and isolation of
CC   full-length gene sequences, including regulatory elements for gene
CC   expression, from genomic human DNA libraries, as hybridisation probes for
CC   screening genomic human DNA libraries. The kits comprising the nucleic
CC   acid sequences are useful for detecting bladder, breast or prostate
CC   cancer cells in a biological sample
XX
SQ   Sequence 135 AA;

  Query Match            100.0%;  Score 694;  DB 4;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-75;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 5
AAY59294
ID   AAY59294 standard; peptide; 21 AA.
XX
AC   AAY59294;
XX
DT   19-APR-2000  (first entry)
XX
DE   Cancer biomarker UC28 antigenic peptide.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Example 5; Page 114; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 21 AA;
```

```
  Query Match             14.8%;  Score 103;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-05;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         54 RKKEKVKRSQKATEFIDYSIE 74
              |||||||||||||||||||||
Db          1 RKKEKVKRSQKATEFIDYSIE 21


RESULT 6
AAU02173
ID   AAU02173 standard; peptide; 21 AA.
XX
AC   AAU02173;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28, antigenic peptide used in diagnosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; antigenic.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Example 5; Col 73; 78pp; English.
XX
CC   The sequence represents the amino acid sequence of biomarker, UC band 28,
CC   antigenic peptide used to produce antibodies for the detection of
CC   prostate, breast or bladder cancer. Biomarker nucleic acid sequences can
CC   be used as hybridisation probes and primers that specifically hybridise
CC   to prostate cancer, benign prostatic hyperplasia (BPH), bladder cancer or
CC   breast cancer markers. The nucleic acids can be used as targets for
CC   therapeutic intervention in these diseases, in the identification and
CC   isolation of full-length gene sequences, including regulatory elements
CC   for gene expression, from genomic human DNA libraries, as hybridisation
CC   probes for screening genomic human DNA libraries. The kits comprising the
CC   nucleic acid sequences are useful for detecting bladder, breast or
CC   prostate cancer cells in a biological sample
XX
SQ   Sequence 21 AA;

  Query Match             14.8%;  Score 103;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-05;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         54 RKKEKVKRSQKATEFIDYSIE 74
              |||||||||||||||||||||
Db          1 RKKEKVKRSQKATEFIDYSIE 21


RESULT 7
ADS41968
ID   ADS41968 standard; protein; 371 AA.
XX
AC   ADS41968;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #20398.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 20398; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 371 AA;

  Query Match             13.0%;  Score 90;  DB 8;  Length 371;
  Best Local Similarity   26.8%;  Pred. No. 0.1;
  Matches   34;  Conservative   23;  Mismatches   38;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRL----SNFP--------------RLPGILAPETVLLPFCYKVF 53
```

```
              |||:   :::|:  | :    :| |               ::|||::|       :
Db         21 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 70

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                :| |::  ||: |: |  ||||  |||      | :| | : |:: :| |
Db         71 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 129

Qy        112 QLTQTLG 118
              :|  |||
Db        130 KLWDTLG 136


RESULT 8
ADN18681
ID   ADN18681 standard; protein; 381 AA.
XX
AC   ADN18681;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #1334.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 1334; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 381 AA;

  Query Match             13.0%;  Score 90;  DB 8;  Length 381;
  Best Local Similarity   26.8%;  Pred. No. 0.11;
  Matches   34;  Conservative   23;  Mismatches   38;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRL----SNFP--------------RLPGILAPETVLLPFCYKVF 53
              |||:   :::|:  | :    :| |               ::|||::|       :
Db         24 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 73

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                :| |::  ||: |: |  ||||  |||      | :| | : |:: :| |
Db         74 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132

Qy        112 QLTQTLG 118
              :|  |||
Db        133 KLWDTLG 139


RESULT 9
ABB90843
ID   ABB90843 standard; protein; 645 AA.
XX
AC   ABB90843;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 54.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP009892.
XX
PR   28-AUG-2001; 2001WO-EP009892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms.
XX
PS   Claim 5; SEQ ID NO 54; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins (ABB90790-ABB94016)
CC   for herbicidally active compounds, comprising aligning and comparing
CC   nucleic acid or amino acid sequences from plant with nucleic acid or
CC   amino acid sequences from non-plant organisms using suitable search
CC   parameters, where plant sequences having an E-value greater by a factor
CC   of 3 than the E-value of most similar non-plant sequences are selected.
CC   The polypeptides or nucleic acids encoding them are useful for
CC   identifying modulators. The identified modulators are useful as
CC   herbicides
XX
SQ   Sequence 645 AA;
```

```
  Query Match             11.9%;  Score 82.5;  DB 5;  Length 645;
  Best Local Similarity   23.5%;  Pred. No. 1.7;
  Matches   31;  Conservative   26;  Mismatches   46;  Indels   29;  Gaps    5;

Qy         25 LRHRLSNF---PRLPGILAPETV-----------------LLPFCYKVFRKKEKVKRSQ 63
              : | |: |   |  | ::   ||                 |||   |   | :   |:
Db         99 MTHLLAGFTYGPHWPWVMTAVTVFKMLTGIVSFLTALSLVTLLPLLLKA--KVREFMLSK 156

Qy         64 KATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKC-SSLSSEAILFTLTLQLTQTLGLECC 122
              |  |     :::    |:   :| | ::  : |  :||    ||:|  ::|::||||: |
Db        157 KTRE-----LDREVGIIMKQTETSLHVRMLTTKIRTSLDRHTILYTTLVELSKTLGLKNC 211

Qy        123 LLYLSKTIHPQI 134
               :::   |  ::
Db        212 AVWIPNEIKTEM 223


RESULT 10
ADN48099
ID   ADN48099 standard; protein; 384 AA.
XX
AC   ADN48099;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Thermococcus kodakaraensis KOD1 protein sequence SeqID1977.
XX
KW   gene disruption; gene targeting; marker gene; transformation;
KW   homologous recombination; hyperthermostable archaebacterium; KOD1;
KW   gene structure; gene function; enzyme activity; medicine;
KW   forensic science; food; drug inspection; molecular biology; immunology.
XX
OS   Thermococcus kodakaraensis.
XX
PN   WO2004022736-A1.
XX
PD   18-MAR-2004.
XX
PF   29-AUG-2003; 2003WO-IB003597.
XX
PR   30-AUG-2002; 2002JP-00319011.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Imanaka T,  Atomi H;
XX
DR   WPI; 2004-257583/24.
XX
PT   Method for disrupting targeted gene in genome of organism particularly
PT   thermostable bacterium and with genome chips for analysis, applicable in
PT   studying gene structure and functions.
XX
PS   Claim 9; SEQ ID NO 1977; 598pp; Japanese.
XX
CC   This invention relates to a novel method for targeting disruption of an
CC   arbitrary gene in a genome of an organism which comprises providing the
CC   whole sequential data of the genome of such organism, selecting at least
CC   1 arbitrary region in the sequence, providing a vector that contains a
CC   sequence homologous with the selected region and a marker gene,
CC   transformation, and homologous recombination. The genome is preferably
CC   the genome of a hyperthermostable archaebacterium, particularly
CC   Thermococcus kodakaraensis KOD1. The method is for targeting the
CC   disruption of a gene in the genome of an organism, which is applicable in
CC   studying gene structure and functions as well as enzyme activities of
CC   encoded proteins and useful in medicine, forensic science, food or drug
CC   inspection, molecular biology and immunology. With this method, the
CC   disruption of a gene at an arbitrary position in a genome can be achieved
CC   efficiently and reliably. The present sequence is that of a protein
CC   encoded by the genome of Thermococcus kodakaraensis which was derived
CC   using the method of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 384 AA;

  Query Match             11.7%;  Score 81;  DB 8;  Length 384;
  Best Local Similarity   23.3%;  Pred. No. 1.3;
  Matches   31;  Conservative   28;  Mismatches   42;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRLS------------------NFPRLPGILAPETVLLPFCYKVF 53
              || :   : ::|| | :|                  ::||:::|  : :   |
Db         21 MHQL--AIYLKKLGHDVSIVTNDLKTGKEKELEELGVGLVKVPGVISP-VLGINITY--- 74

Qy         54 RKKEKVKRSQKATEF-IDYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                    :| :::  || :|: :  :||| |||        | :: | : |:: :| |:
Db         75 ----GLKSNRELGEFLVDFDVVHAHHA-FTPLSLKAVKAGRTLEKATLLTTHSISFSHES 129

Qy        112 QLTQTLGLECCLL 124
               | : |||   ||
Db        130 SLWKALGLTFPLL 142


RESULT 11
ADF50279
ID   ADF50279 standard; protein; 1061 AA.
XX
AC   ADF50279;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human PFM7 protein, a PR/SET family member.
XX
KW   human; PFM7; SET; Su(var)3-9, Enhancer-of-Zeste and Trithorax; cancer;
KW   PFM/SET; tumour suppressor; hyperproliferative disorder; scleroderma;
KW   arthritis; alcoholic liver cirrhosis; hypertropic scarring;
KW   atherosclerosis; gene therapy; Duchenne's muscular dystrophy;
KW   insulin-dependent diabetes mellitus; Huntington's; Parkinson's;
KW   Alzheimer's disease; paralysis; cerebellar atrophy; cytostatic;
KW   neuroprotective; nootropic; antiarthritic; vulnerary;
KW   antiatherosclerotic; antidiabetic; hepatotropic; PFM; PR family member.
XX
OS   Homo sapiens.
XX
PN   US2003049623-A1.
XX
PD   13-MAR-2003.
XX
PF   18-JUL-2001; 2001US-00910478.
XX
PR   18-JUL-2001; 2001US-00910478.
XX
PA   (HUAN/) HUANG S.
XX
PI   Huang S;
XX
DR   WPI; 2003-567062/53.
DR   N-PSDB; ADF50278.
XX
PT   New PR/SET-domain containing nucleic acids (which encodes PFM/SET) and
PT   polypeptides, useful for preventing or treating cancers, scleroderma,
PT   arthritis, keloids, atherosclerosis, Huntington's disease or Alzheimer's
PT   disease.
XX
PS   Claim 21; SEQ ID NO 4; 83pp; English.
XX
CC   This invention relates to novel PFM (PR family member)/SET (Su(var)3-9,
CC   Enhancer-of-Zeste and Trithorax) nucleotides and encoded polypeptides
CC   thereof that are implicated in proliferative disorders such as cancer.
CC   Specifically, it refers to a functional fragment of the PFM/SET tumour
CC   suppressor gene that encodes a PR, SET, PRAZ or PKZL domain. The present
CC   invention describes the identification and characterisation of additional
CC   PR/SET-domain family members that can be used as regulators of cell
CC   proliferation and furthermore to treat, prevent or diagnose
CC   hyperproliferative disorders including scleroderma, arthritis, alcoholic
```

```
CC   liver cirrhosis, hypertropic scarring and atherosclerosis. Through gene
CC   therapy, these polynucleotides can be used to enhance proliferation of
CC   normal cells without rendering the cells cancerous and as such they are
CC   particularly useful for treating Duchenne's muscular dystrophy, insulin-
CC   dependent diabetes mellitus, Huntington's, Parkinson's, Alzheimer's
CC   disease, paralysis, or cerebellar atrophy. Accordingly, these
CC   compositions can be described as cytostatic, neuroprotective, nootropic,
CC   antiarthritic, vulnerary, antiatherosclerotic, antidiabetic or
CC   hepatotropic. This polypeptide sequence is the human PFM7 protein
CC   (encoded by a PFM/SET gene localised to chromosome 11q25) of the
CC   invention.
XX
SQ   Sequence 1061 AA;

  Query Match             11.2%;  Score 77.5;  DB 7;  Length 1061;
  Best Local Similarity   28.6%;  Pred. No. 13;
  Matches   36;  Conservative   17;  Mismatches   42;  Indels   31;  Gaps    6;

Qy         18 HILQ----------IRKLR--------HRLSNFPRLPGILAPETVLLPFCYKVFRKKEKV 59
              |||:          |||||          ||   :| | :|   |  | | | :  | |:
Db        704 HILKNHPGAELPPSIRKLRPAGPGEPDPMLSTHTQLTGTIATPPVCCPHCSKQYSSKTKM 763

Qy         60 KR--SQKATEFIDYSIEQSHHAILTPLQTH-LTMKGSSMKCSSLSSEAILFTLTL----- 111
               :   :|  ||   |     : | ||| |  ::   : :   | : | :: |  |
Db        764 VQHIRKKHPEFAQLS-----NTIHTPLTTAVISATPAVLTTDSATGETVVTTDLLTQAMT 818

Qy        112 QLTQTL 117
              :|:|||
Db        819 ELSQTL 824


RESULT 12
ADO36741
ID   ADO36741 standard; protein; 1061 AA.
XX
AC   ADO36741;
XX
DT   15-JUL-2004  (first entry)
XX
DE   Human PR-domain containing protein PFM-7.
XX
KW   cytostatic; histone methyltransferase; gene therapy;
KW   PFM/SET protein binding motif; cell growth modulator;
KW   histone methyltransferase activity; proliferative disorder; cancer;
KW   PR-domain; PFM-7.
XX
OS   Homo sapiens.
XX
PN   US2004014192-A1.
XX
PD   22-JAN-2004.
XX
PF   18-JUL-2002; 2002US-00200012.
XX
PR   18-JUL-2002; 2002US-00200012.
XX
PA   (HUAN/) HUANG S.
XX
PI   Huang S;
XX
DR   WPI; 2004-121568/12.
XX
PT   New isolated nucleic acid molecule comprising a sequence encoding a
PT   PFM/SET polypeptide, useful for diagnosing, prognosing, preventing and
PT   treating proliferative disorders, e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 4; 38pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule comprising a
CC   sequence encoding a PFM/SET protein binding motifs polypeptides
CC   comprising a 689, 1061, 367, 717, 571, 951, 1257, 720, or 770 amino acid
CC   sequence (P1-P9), given in the specification. Also described are: a
CC   vector comprising the isolated nucleic acid molecule; a host cell
CC   comprising the vector; an oligonucleotide comprising at least 17
CC   contiguous nucleotides of the nucleic acid molecule; a primer pair for
CC   detecting PFM/SET nucleic acid molecule, comprising two isolated
CC   oligonucleotides; detecting PFM/SET nucleic acid molecule in a sample;
CC   modulating cell growth by introducing the vector into a host cell, and
CC   expressing the encoded PFM/SET polypeptide in an amount effective to
CC   modulate growth of the cell; isolating PFM/SET polypeptide by growing the
CC   host cell under conditions appropriate for the expression of the
CC   polypeptide; the isolated PFM/SET polypeptide, or its functional fragment
CC   ; an isolated immunogenic PFM/SET peptide, comprising at least 8
CC   contiguous amino acids of P1-P9; an antibody or its antigen-binding
CC   fragment that specifically binds to PFM/SET polypeptide; detecting
CC   PFM/SET polypeptide in a sample; and screening for a compound that
CC   modulates PFM/SET polypeptide histone methyltransferase activity by
CC   contacting the polypeptide with one or more candidate compounds, and
CC   determining histone methyltransferase activity of the contacted
CC   polypeptide. The nucleic acid molecules, polypeptides and methods are
CC   useful for diagnosing, prognosing, preventing and treating proliferative
CC   disorders, e.g. cancer. This is the amino acid sequence of PR-domain
CC   containing polypeptide PFM-7.
XX
SQ   Sequence 1061 AA;

  Query Match             11.2%;  Score 77.5;  DB 8;  Length 1061;
  Best Local Similarity   28.6%;  Pred. No. 13;
  Matches   36;  Conservative   17;  Mismatches   42;  Indels   31;  Gaps    6;

Qy         18 HILQ----------IRKLR--------HRLSNFPRLPGILAPETVLLPFCYKVFRKKEKV 59
              |||:          |||||          ||   :| | :|   |  | | | :  | |:
Db        704 HILKNHPGAELPPSIRKLRPAGPGEPDPMLSTHTQLTGTIATPPVCCPHCSKQYSSKTKM 763

Qy         60 KR--SQKATEFIDYSIEQSHHAILTPLQTH-LTMKGSSMKCSSLSSEAILFTLTL----- 111
               :   :|  ||   |     : | ||| |  ::   : :   | : | :: |  |
Db        764 VQHIRKKHPEFAQLS-----NTIHTPLTTAVISATPAVLTTDSATGETVVTTDLLTQAMT 818

Qy        112 QLTQTL 117
              :|:|||
Db        819 ELSQTL 824


RESULT 13
ABB58124
ID   ABB58124 standard; protein; 1144 AA.
XX
AC   ABB58124;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 1164.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02227.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
```

```
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 1164; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1144 AA;

  Query Match             11.2%;  Score 77.5;  DB 4;  Length 1144;
  Best Local Similarity   30.0%;  Pred. No. 15;
  Matches   30;  Conservative   12;  Mismatches   33;  Indels   25;  Gaps    5;

Qy         15 NIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKK------EKVKRSQKATEF 68
              | :| || |||||  ||          : : ||  :| |  |      || |   :  :|
Db        718 NQVHELQ-RKLRHVSSNL---------KAIDLPTNFKFFFLKTDRHGQEKAK--SQIQKF 765

Qy         69 IDYSIEQSH-------HAILTPLQTHLTMKGSSMKCSSLS 101
              ::: :|  |       :  |:|  ||     | | |  |
Db        766 LNFILEDDHLNGSEAIYTFLSPSSDHLKQSLPSPKKSKFS 805


RESULT 14
ADS08184
ID   ADS08184 standard; protein; 1026 AA.
XX
AC   ADS08184;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Staphylococcus epidermis polypeptide seqid 7479.
XX
KW   antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW   recombinant expression vector; infection; computer readable medium;
KW   computer based system.
XX
OS   Staphylococcus epidermidis.
XX
PN   US2004147734-A1.
XX
PD   29-JUL-2004.
XX
PF   01-DEC-2003; 2003US-00724972.
XX
PR   08-NOV-1997;   97US-0064964P.
PR   13-AUG-1998;   98US-00134001.
PR   29-NOV-1999;   99US-00450969.
XX
PA   (DOUC/) DOUCETTE-STAMM L.
PA   (BUSH/) BUSH D.
XX
PI   Doucette-Stamm L,  Bush D;
XX
DR   WPI; 2004-580138/56.
DR   N-PSDB; ADS04412.
XX
PT   New isolated polypeptide and encoding nucleic acid derived from
PT   Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT   treating an S. epidermidis bacterial infection.
XX
PS   Claim 17; SEQ ID NO 7479; 741pp; English.
XX
CC   The invention describes an isolated nucleic acid comprising a nucleotide
CC   sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC   1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC   of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC   given in the specification. Also described are: a recombinant expression
CC   vector; a cell comprising a recombinant expression vector of (1);
CC   producing an S. epidermidis polypeptide; an isolated nucleic acid
CC   comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC   vaccine composition for prevention or treatment of an S. epidermidis
CC   infection, comprising a nucleic acid cited above and a carrier; treating
CC   a subject for S. epidermidis infection; a recombinant or substantially
CC   pure preparation of an S. epidermidis polypeptide or its fragment; a
CC   vaccine composition for prevention or treatment of an S. epidermidis
CC   infection; detecting the presence of a Staphylococcus nucleic acid in a
CC   sample; a computer readable medium having recorded in it the nucleotide
CC   sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC   system for identifying fragments of the Staphylococcus genome of
CC   commercial importance; a computer based system for identifying fragments
CC   of the Staphylococcus plasmids of commercial importance; identifying
CC   commercially important nucleic acid fragments of the Staphylococcus
CC   genome and/or plasmids; and identifying an expression modulating fragment
CC   of the Staphylococcus genome and/or plasmids. The methods and
CC   compositions of the present invention are useful for the diagnosis,
CC   prevention and/or treatment of an Staphylococcal epidermidis bacterial
CC   infection. This is the amino acid sequence of a S. epidermis protein of
CC   the invention.
XX
SQ   Sequence 1026 AA;

  Query Match             11.1%;  Score 77;  DB 8;  Length 1026;
  Best Local Similarity   20.6%;  Pred. No. 15;
  Matches   35;  Conservative   33;  Mismatches   54;  Indels   48;  Gaps    8;

Qy          7 NQKYED-MHNIIHILQIR---------KLRHRLSNFPRLPGILAPETVL----LPFCYKV 52
              ||  || || :| |:::           |::  ||       :: | : |    ||  | :
Db        782 NQALEDEMHTVIKIVELSRQARKNADLKIKQPLSKM-----VIKPNSQLNLSFLPNYYSI 836

Qy         53 FRKKEKVKR---SQKATEFIDYSIEQSHHAI-----------------LTPLQTHLTMKG 92
                : :  :|    :    ::| | :: :  ::                 |:       ::
Db        837 IKDELNIKNIELTDNINDYITYELKLNFSSVGPKLGNKTKNIQTLIDSLSEYDKKSLIES 896

Qy         93 SSMKCSSLSSEA-------ILFTLTLQLTQTLGLECCLLYLSKTIHPQII 135
              :: |  ||||:|       |: ||      |       |:: | | : |::|
Db        897 NNFK--SLSSDAELTKDDFIIKTLPKDSYQLSEDNDCVILLDKNLSPELI 944


RESULT 15
ABM69043
ID   ABM69043 standard; protein; 322 AA.
XX
AC   ABM69043;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Photorhabdus luminescens protein sequence #2140.
XX
KW   Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW   detection; food; gene expression; plant; animal; microorganism; toxin;
KW   antibiotic; biopesticide; virulence factor; disease model; plague;
KW   whooping cough.
XX
OS   Photorhabdus luminescens.
XX
PN   WO200294867-A2.
XX
PD   28-NOV-2002.
XX
PF   07-FEB-2002; 2002WO-IB003040.
XX
PR   07-FEB-2001; 2001FR-00001659.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Duchaud E,  Taourit S,  Glaser P,  Frangeul L,  Kunst F,  Danchin A;
PI   Buchrieser C;
```

XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 2140; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 322 AA;

Query Match 11.0%; Score 76.5; DB 6; Length 322;
Best Local Similarity 26.6%; Pred. No. 3.5;
Matches 29; Conservative 19; Mismatches 50; Indels 11; Gaps 3;

```
Qy         13 MHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVKRSQKATEFIDYS 72
              :|| || :||   :  ||    ||||      ::    :|          ||| | | :
Db        221 IHNFIHAVQIPAEKFGLSRTVNLPGISVTVQEMIDALAEV--------AGQKAVELIRFE 272

Qy         73 IEQSHHAILTPLQTHLTM-KGSSM--KCSSLSSEAILFTLTLQLTQTLG 118
               ::: : |:     |  : :| |:     :  |: |   :|  |:|| |
Db        273 PDENINRIVASWPGHFDISRGLSLGFHADNTFSDTIRAFITNNLSQTGG 321
```

Search completed: September 9, 2005, 13:00:51
Job time : 62.5 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:25:33 ; Search time 17.5 Seconds
(without alignments)
575.864 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	3	US-09-097-199-84	Sequence 84, Appl
2	694	100.0	135	3	US-09-097-199-86	Sequence 86, Appl
3	694	100.0	157	4	US-09-949-016-9182	Sequence 9182, Ap
4	103	14.8	21	3	US-09-097-199-56	Sequence 56, Appl
5	72	10.4	636	4	US-09-614-912-92	Sequence 92, Appl
6	71.5	10.3	634	4	US-09-614-912-88	Sequence 88, Appl
7	70.5	10.2	438	4	US-09-489-039A-11150	Sequence 11150, A
8	70.5	10.2	634	4	US-09-614-912-100	Sequence 100, App
9	69.5	10.0	1294	3	US-09-473-717-2	Sequence 2, Appli
10	69.5	10.0	1294	4	US-09-949-016-6023	Sequence 6023, Ap
11	69.5	10.0	1353	3	US-09-398-193-99	Sequence 99, Appl
12	68.5	9.9	2470	4	US-08-265-967C-2	Sequence 2, Appli
13	68.5	9.9	2470	4	US-08-305-790B-3	Sequence 3, Appli
14	68	9.8	496	4	US-10-146-704-3	Sequence 3, Appli
15	68	9.8	639	4	US-09-949-016-6812	Sequence 6812, Ap
16	68	9.8	652	4	US-09-949-016-7323	Sequence 7323, Ap
17	68	9.8	993	4	US-09-949-016-10335	Sequence 10335, A
18	67.5	9.7	156	4	US-09-543-681A-7434	Sequence 7434, Ap
19	67	9.7	399	4	US-09-543-681A-4933	Sequence 4933, Ap
20	67	9.7	517	3	US-09-457-040B-14	Sequence 14, Appl
21	67	9.7	773	2	US-08-484-101B-42	Sequence 42, Appl
22	67	9.7	773	2	US-08-484-101B-44	Sequence 44, Appl
23	67	9.7	773	3	US-08-714-524D-42	Sequence 42, Appl
24	67	9.7	773	3	US-08-714-524D-44	Sequence 44, Appl
25	66.5	9.6	580	4	US-09-107-532A-5734	Sequence 5734, Ap
26	66.5	9.6	1784	3	US-09-040-738-2	Sequence 2, Appli
27	66.5	9.6	1784	3	US-08-652-426A-2	Sequence 2, Appli
28	66.5	9.6	1813	4	US-09-949-016-8283	Sequence 8283, Ap
29	66.5	9.6	1813	4	US-09-949-016-8284	Sequence 8284, Ap
30	66.5	9.6	1813	4	US-09-949-016-8285	Sequence 8285, Ap
31	66	9.5	267	4	US-09-270-767-46386	Sequence 46386, A
32	66	9.5	348	4	US-09-198-452A-922	Sequence 922, App
33	66	9.5	348	4	US-09-438-185A-856	Sequence 856, App
34	66	9.5	611	4	US-09-949-016-7139	Sequence 7139, Ap
35	65.5	9.4	698	4	US-09-538-092-151	Sequence 151, App
36	65.5	9.4	3218	1	US-08-764-100-27	Sequence 27, Appl
37	65	9.4	75	4	US-09-621-976-6155	Sequence 6155, Ap
38	65	9.4	258	4	US-09-270-767-33067	Sequence 33067, A
39	65	9.4	258	4	US-09-270-767-48284	Sequence 48284, A
40	65	9.4	288	1	US-08-375-709-9	Sequence 9, Appli
41	65	9.4	288	1	US-08-752-929-9	Sequence 9, Appli
42	65	9.4	288	3	US-09-090-793-6	Sequence 6, Appli
43	65	9.4	288	4	US-09-231-899-6	Sequence 6, Appli
44	64.5	9.3	236	4	US-09-107-532A-6873	Sequence 6873, Ap
45	64.5	9.3	315	3	US-09-134-001C-5446	Sequence 5446, Ap

ALIGNMENTS

```
RESULT 1
US-09-097-199-84
; Sequence 84, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  84:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-097-199-84

  Query Match           100.0%;  Score 694;  DB 3;  Length 135;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-77;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 2
US-09-097-199-86
; Sequence 86, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  86:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-097-199-86

  Query Match             100.0%;  Score 694;  DB 3;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-77;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 3
US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9182
;   LENGTH: 157
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-9182

  Query Match             100.0%;  Score 694;  DB 4;  Length 157;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-77;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 82

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 142

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        143 CCLLYLSKTIHPQII 157


RESULT 4
US-09-097-199-56
; Sequence 56, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:59:09 ; Search time 52.5 Seconds
(without alignments)
1014.252 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 1777461

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	10	US-09-974-546-84	Sequence 84, Appl
2	694	100.0	135	10	US-09-974-546-86	Sequence 86, Appl
3	103	14.8	21	10	US-09-974-546-56	Sequence 56, Appl
4	90	13.0	371	15	US-10-369-493-20398	Sequence 20398, A
5	90	13.0	381	15	US-10-369-493-1334	Sequence 1334, Ap
6	81	11.7	1478	16	US-10-425-115-362997	Sequence 362997,
7	79	11.4	264	15	US-10-424-599-174106	Sequence 174106,
8	79	11.4	1368	16	US-10-425-115-324108	Sequence 324108,
9	79	11.4	1444	16	US-10-425-115-324095	Sequence 324095,
10	78	11.2	289	15	US-10-424-599-148597	Sequence 148597,
11	78	11.2	3229	16	US-10-425-115-324093	Sequence 324093,
12	77.5	11.2	1061	15	US-10-200-012-4	Sequence 4, Appli
13	77	11.1	1026	18	US-10-724-972A-7479	Sequence 7479, Ap
14	77	11.1	1766	16	US-10-425-115-324098	Sequence 324098,
15	75	10.8	633	15	US-10-282-122A-70203	Sequence 70203, A
16	75	10.8	1259	16	US-10-425-115-324122	Sequence 324122,
17	75	10.8	3749	16	US-10-425-115-324125	Sequence 324125,
18	74.5	10.7	388	15	US-10-424-599-169767	Sequence 169767,
19	74.5	10.7	404	18	US-10-954-778-97	Sequence 97, Appl
20	74	10.7	1353	16	US-10-425-115-262744	Sequence 262744,
21	74	10.7	1605	16	US-10-425-115-362998	Sequence 362998,
22	74	10.7	1676	16	US-10-425-115-238569	Sequence 238569,
23	74	10.7	1708	16	US-10-425-115-324103	Sequence 324103,
24	74	10.7	1752	16	US-10-425-115-238562	Sequence 238562,
25	74	10.7	1754	16	US-10-425-115-324104	Sequence 324104,
26	73	10.5	1143	16	US-10-425-115-238564	Sequence 238564,
27	73	10.5	1517	16	US-10-425-115-324099	Sequence 324099,
28	72.5	10.4	3365	16	US-10-739-930-6537	Sequence 6537, Ap
29	72	10.4	329	15	US-10-424-599-150082	Sequence 150082,
30	71.5	10.3	407	15	US-10-425-114-69882	Sequence 69882, A
31	71.5	10.3	426	16	US-10-425-115-327356	Sequence 327356,
32	71.5	10.3	676	15	US-10-424-599-267862	Sequence 267862,
33	71	10.2	334	15	US-10-425-114-63742	Sequence 63742, A
34	71	10.2	1214	16	US-10-425-115-243074	Sequence 243074,
35	71	10.2	1595	16	US-10-425-115-238559	Sequence 238559,
36	71	10.2	1715	16	US-10-425-115-324123	Sequence 324123,
37	71	10.2	1924	9	US-09-866-557A-2	Sequence 2, Appli
38	71	10.2	1924	11	US-09-858-862-2	Sequence 2, Appli
39	71	10.2	1924	14	US-10-055-797-2	Sequence 2, Appli
40	71	10.2	1924	15	US-10-350-798-2	Sequence 2, Appli
41	70.5	10.2	269	15	US-10-424-599-264790	Sequence 264790,
42	70.5	10.2	634	17	US-10-876-086-27	Sequence 27, Appl
43	70	10.1	206	16	US-10-408-765A-1698	Sequence 1698, Ap
44	70	10.1	272	16	US-10-437-963-125982	Sequence 125982,
45	70	10.1	327	15	US-10-425-114-63743	Sequence 63743, A

ALIGNMENTS

RESULT 1
US-09-974-546-84
; Sequence 84, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199
; FILING DATE: 1998-06-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023

```
;             REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 84:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 135 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

  Query Match             100.0%;  Score 694;  DB 10;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-72;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 2
US-09-974-546-86
; Sequence 86, Application US/09974546
; Publication No. US20030050470A1
;   GENERAL INFORMATION:
;         APPLICANT: An, Gang
;                    O'Hara, S. Mark
;                    Ralph, David
;                    Veltri, Robert
;         TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                             PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;         NUMBER OF SEQUENCES: 87
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/974,546
;             FILING DATE: 10-Oct-2001
;             CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/097,199
;             FILING DATE: 1998-06-12
;         ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 86:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 135 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

  Query Match             100.0%;  Score 694;  DB 10;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-72;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 3
US-09-974-546-56
; Sequence 56, Application US/09974546
; Publication No. US20030050470A1
;   GENERAL INFORMATION:
;         APPLICANT: An, Gang
;                    O'Hara, S. Mark
;                    Ralph, David
;                    Veltri, Robert
;         TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                             PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;         NUMBER OF SEQUENCES: 87
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/974,546
;             FILING DATE: 10-Oct-2001
;             CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/097,199
;             FILING DATE: 1998-06-12
;         ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 56:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 21 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-974-546-56

  Query Match             14.8%;  Score 103;  DB 10;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         54 RKKEKVKRSQKATEFIDYSIE 74
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:01:39 ; Search time 14.5 Seconds
(without alignments)
895.811 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90	13.0	381	2	F71196	probable hexosyltr
2	82.5	11.9	645	2	F86174	hypothetical prote
3	80	11.5	495	1	MNXRSA	nonstructural prot
4	78.5	11.3	3147	2	T21328	hypothetical prote
5	78	11.2	523	2	D64555	conserved hypothet
6	74.5	10.7	172	2	A58456	X protein - human
7	74.5	10.7	263	2	A82069	hypothetical prote
8	72.5	10.4	796	2	T43782	hypothetical prote
9	71.5	10.3	227	2	B84040	ABC transporter (A
10	71.5	10.3	451	2	D70045	two-component sens
11	70.5	10.2	1064	2	T40751	isoleucyl-trna syn
12	70	10.1	404	2	T06761	hypothetical prote
13	70	10.1	982	1	A44831	phosphoenolpyruvat
14	70	10.1	982	2	AE2413	phosphoenolpyruvat
15	69.5	10.0	1398	2	H71606	hypothetical prote
16	69	9.9	154	2	T27983	hypothetical prote
17	69	9.9	743	2	T00634	hypothetical prote
18	69	9.9	1850	2	AC1917	serine/threonine k
19	68.5	9.9	503	1	S11338	steroid 11beta-mon
20	68.5	9.9	2470	2	S57085	1-phosphatidylinos
21	68	9.8	379	2	C75006	hypothetical prote
22	68	9.8	414	1	H64203	histidine-tRNA lig
23	68	9.8	496	2	A31986	glucose transporte
24	68	9.8	639	2	T13151	adapter protein CM
25	68	9.8	5126	2	S40450	ryanodine receptor
26	67.5	9.7	619	2	B87682	hypothetical prote
27	67.5	9.7	833	2	T28385	ORF MSV224 probabl
28	67.5	9.7	1278	2	A71609	probable secreted
29	67.5	9.7	3225	2	D81702	adherence factor T
30	67	9.7	126	2	T03757	hypothetical prote
31	67	9.7	213	2	E96506	hypothetical prote
32	67	9.7	317	2	B82440	hypothetical prote
33	67	9.7	347	2	T01452	translation initia
34	67	9.7	517	2	A54099	protein kinase Dar
35	67	9.7	587	2	S37927	hypothetical prote
36	66.5	9.6	244	2	AC3341	precorrin-2 C20-me
37	66.5	9.6	447	2	T21716	hypothetical prote
38	66.5	9.6	635	2	T06537	ethylene receptor
39	66.5	9.6	859	2	H70327	DNA mismatch repai
40	66.5	9.6	908	2	T22376	hypothetical prote
41	66.5	9.6	1784	2	A49420	tuberous sclerosis
42	66	9.5	344	2	D86597	outer membrane pro
43	66	9.5	344	2	H72027	major outer membra
44	66	9.5	446	2	A42029	transcription fact
45	66	9.5	472	2	F97147	probable Fe-S oxid

ALIGNMENTS

RESULT 1
F71196
probable hexosyltransferase (EC 2.4.1.-) PH1844 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: F71196
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71196
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-381 <KAW>
A;Cross-references: UNIPROT:O59512; GB:AP000007; NID:g3236134; PIDN:BAA30965.1; PID:g325
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1844
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 13.0%; Score 90; DB 2; Length 381;
Best Local Similarity 26.8%; Pred. No. 0.16;
Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;

```
Qy         13 MHNIIHILQIRKLRHRL----SNFP---------------RLPGILAPETVLLPFCYKVF 53
              |||:   :::|:  | :    :| |               ::|||::|        :
Db         24 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 73

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                 :| |::  ||: |: |  |||| |||       | :| | : |:: :| |
Db         74 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132

Qy        112 QLTQTLG 118
              :|  |||
Db        133 KLWDTLG 139
```

RESULT 2
F86174
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: F86174
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-645 <STO>
A;Cross-references: UNIPROT:P93825; GB:AE005172; NID:g2341041; PIDN:AAB70445.1; GSPDB:GN
C;Genetics:
A;Map position: 1

Query Match 11.9%; Score 82.5; DB 2; Length 645;
Best Local Similarity 23.5%; Pred. No. 1.7;
Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;

```
Qy         25 LRHRLSNF---PRLPGILAPETV-----------------LLPFCYKVFRKKEKVKRSQ 63
              : | |: |   |  | ::   ||                  |||   |   |  :   |:
Db         99 MTHLLAGFTYGPHWPWVMTAVTVFKMLTGIVSFLTALSLVTLLPLLLKA--KVREFMLSK 156

Qy         64 KATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKC-SSLSSEAILFTLTLQLTQTLGLECC 122
              |  |     :::   |:   :| | ::  : |  :||    ||:|  ::|::||||: |
Db        157 KTRE-----LDREVGIIMKQTETSLHVRMLTTKIRTSLDRHTILYTTLVELSKTLGLKNC 211

Qy        123 LLYLSKTIHPQI 134
               :::   |  ::
Db        212 AVWIPNEIKTEM 223
```

RESULT 3
MNXRSA
nonstructural protein NCVP2 - simian rotavirus SA11
N;Alternate names: nonstructural protein NS53
C;Species: simian rotavirus SA11
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S08215
R;Mitchell, D.B.; Both, G.W.
Virology 174, 618-621, 1990
A;Title: Conservation of a potential metal binding motif despite extensive sequence dive
A;Reference number: S08215; MUID:90163231; PMID:2154894
A;Accession: S08215
A;Molecule type: genomic RNA
A;Residues: 1-495 <MIT>
A;Cross-references: UNIPROT:P15687; EMBL:X14914; NID:g61889; PIDN:CAA33039.1; PID:g61890
C;Genetics:
A;Map position: segment 5
C;Superfamily: bovine rotavirus nonstructural protein NCVP2
C;Keywords: nonstructural protein; zinc finger

Query Match 11.5%; Score 80; DB 1; Length 495;
Best Local Similarity 25.3%; Pred. No. 2.3;
Matches 24; Conservative 22; Mismatches 33; Indels 16; Gaps 4;

```
Qy          2 RAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVKR 61
              | || |:          |:|::| ::| ::   :|  |:    :: |   || || |:: :
Db         71 RCFLDNEP--------HLLKLRTVKHPITK-DKLQCIIDLYNIIFPINDKVIRKFERMIK 121

Qy         62 SQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMK 96
              :|      | ||  :| :|      :|:  :: |
Db        122 QRKCRN--QYKIEWYNHLLLP-----ITLNAAAFK 149
```

RESULT 4
T21328
hypothetical protein F25C8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21328
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19407
A;Accession: T21328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3147 <WIL>
A;Cross-references: UNIPROT:Q9XV66; EMBL:Z81512; PIDN:CAB04172.1; GSPDB:GN00023; CESP:F25
A;Experimental source: clone F25C8
C;Genetics:
A;Gene: CESP:F25C8.3
A;Map position: 5
A;Introns: 32/2; 57/2; 73/3; 113/3; 157/3; 366/1; 456/1; 509/1; 638/1; 773/2; 848/1; 902/
/3; 2510/2; 2737/3; 2827/1; 2902/3; 3093/3; 3121/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F25C8.3

Query Match 11.3%; Score 78.5; DB 2; Length 3147;
Best Local Similarity 28.1%; Pred. No. 26;
Matches 41; Conservative 17; Mismatches 67; Indels 21; Gaps 6;

```
Qy          2 RAFLRN-------QKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFR 54
              ||  ||         | : :   ||  |   :  | |      |   | : : :    |: :
Db       1571 RADKRNLLADNWQAKQQALRKSIHARQSTAVPRRESAMVGQPE-FASKAIKMMLMEKMQQ 1629

Qy         55 KKEKVKRSQKATE--FIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQ 112
              :||| |  :|  :       |:|| | :  |      |  |   |           | :| ||
Db       1630 EKEKEKEKEKEEKDALKKQSVEQDHSSTDTEEDAQLPEKNKPML-------TYLRSLVLQ 1682

Qy        113 LTQ---TLGLECCLLYLSKTIHPQII 135
              |     :  |:|||| ||   | |:|
Db       1683 LVHSPISSVLKCCLL-LSVEQHKQMI 1707
```

RESULT 5
D64555
conserved hypothetical integral membrane protein HP0284 - Helicobacter pylori (strain 26
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: D64555
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: D64555
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-523 <TOM>
A;Cross-references: UNIPROT:O25059; GB:AE000547; GB:AE000511; NID:g2313377; PIDN:AAD07352

Query Match 11.2%; Score 78; DB 2; Length 523;
Best Local Similarity 23.5%; Pred. No. 3.9;
Matches 24; Conservative 16; Mismatches 30; Indels 32; Gaps 3;

```
Qy          5 LRNQKYEDMHNIIHILQIRKLRHRLS----------------------NFPRLPGILA 40
              |:|:| :|      :|:|  | | |                      | | :| |
Db         71 LKNKKSKD------LLKISTLEHTLKALESQQKMFESYGVNPFKDLIERPNIPNIPNIAN 124

Qy         41 PETVL--LPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAI 80
              |  ::  : |   :  | | :||:| | | :   ::| |  :
Db        125 PIAIIDGISFIKSMHLKHENLKRNQTALEEVLRLLDQKHQLL 166
```

RESULT 6
A58456
X protein - human hepatitis B virus
C;Species: human hepatitis B virus
C;Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 09-Jul-2004
C;Accession: A58456
R;Zeng, M.; Huang, B.R.; Cai, L.W.; Pan, G.Z.
Chinese Biochem. J. 12, 22-26, 1996
A;Title: The sequence analysis of the hepatitis B virus (HBV) X gene and its expression i
A;Reference number: JC1331
A;Accession: A58456
A;Molecule type: DNA
A;Residues: 1-172 <ZEN>

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:00:59 ; Search time 51.5 Seconds
(without alignments)
1342.343 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	2	Q9GZY1	Q9gzy1 homo sapien
2	90	13.0	381	2	O59512	O59512 pyrococcus
3	82.5	11.9	645	2	P93825	P93825 arabidopsis
4	81	11.7	451	2	Q7SZX8	Q7szx8 brachydanio
5	80	11.5	414	2	Q99FX6	Q99fx6 simian rota
6	80	11.5	479	2	Q99FX4	Q99fx4 simian rota
7	80	11.5	495	1	VN53_ROTS1	P15687 simian 11 r
8	80	11.5	496	2	Q99FX5	Q99fx5 simian rota
9	80	11.5	496	2	Q99FX7	Q99fx7 simian rota
10	78.5	11.3	3175	2	Q7JKT8	Q7jkt8 caenorhabdi
11	78	11.2	523	2	O25059	O25059 helicobacte
12	78	11.2	1646	2	Q86820	Q86820 potato aucu
13	77.5	11.2	358	2	Q8TZU8	Q8tzu8 pyrococcus
14	77.5	11.2	931	2	Q9NXI4	Q9nxi4 homo sapien
15	77.5	11.2	1023	2	Q863Z2	Q863z2 homo sapien
16	77.5	11.2	1061	1	PRDA_HUMAN	Q9nqv6 homo sapien
17	77.5	11.2	1061	2	Q9ULI9	Q9uli9 homo sapien
18	77.5	11.2	1117	2	Q8T4E8	Q8t4e8 drosophila
19	77.5	11.2	1117	2	Q9W3N0	Q9w3n0 drosophila
20	77	11.1	495	1	VN53_ROTSP	P35425 simian 11 r
21	77	11.1	1024	1	SYIP_STAAU	P41368 staphylococ
22	76.5	11.0	321	2	Q7N442	Q7n442 photorhabdu
23	76.5	11.0	398	2	Q6LRR8	Q6lrr8 photobacter
24	76.5	11.0	2259	1	YCF2_PHYPA	P61243 physcomitre
25	75	10.8	629	2	Q6GJ08	Q6gj08 staphylococ
26	74.5	10.7	160	2	Q6SIF8	Q6sif8 uncultured
27	74.5	10.7	163	2	Q6SCJ3	Q6scj3 uncultured
28	74.5	10.7	172	2	Q7M022	Q7m022 hepatitis b
29	74.5	10.7	263	2	Q9KP79	Q9kp79 vibrio chol
30	74.5	10.7	404	2	Q7PNX8	Q7pnx8 anopheles g
31	74	10.7	3184	2	Q9XV66	Q9xv66 caenorhabdi
32	73.5	10.6	615	2	Q9FHK4	Q9fhk4 arabidopsis
33	73.5	10.6	853	2	Q7R2I5	Q7r2i5 giardia lam
34	72.5	10.4	220	2	Q7P537	Q7p537 fusobacteri
35	72.5	10.4	220	2	Q8RF89	Q8rf89 fusobacteri
36	72.5	10.4	716	2	O21040	O21040 dictyosteli
37	72.5	10.4	796	2	Q9TGM3	Q9tgm3 dictyosteli
38	72.5	10.4	1283	2	O97034	O97034 ephydatia f
39	72.5	10.4	3306	2	Q9FT44	Q9ft44 arabidopsis
40	72	10.4	316	2	Q83DP2	Q83dp2 coxiella bu
41	72	10.4	400	2	Q6RX30	Q6rx30 arabidopsis
42	72	10.4	448	2	Q6ALV1	Q6alv1 desulfotale
43	72	10.4	475	2	Q8XLZ4	Q8xlz4 clostridium
44	72	10.4	788	2	Q7RJY2	Q7rjy2 plasmodium
45	71.5	10.3	227	2	Q9K884	Q9k884 bacillus ha

ALIGNMENTS

RESULT 1
Q9GZY1
ID Q9GZY1 PRELIMINARY; PRT; 135 AA.
AC Q9GZY1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RA Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1; -.
DR EMBL; AF189269; AAG17117.1; -.
DR Genew; HGNC:21079; PBOV1.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B1983705D0 CRC64;

Query Match 100.0%; Score 694; DB 2; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.9e-60;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135
```

RESULT 2
O59512
ID O59512 PRELIMINARY; PRT; 381 AA.
AC O59512;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PH1844.
GN OrderedLocusNames=PH1844;
OS Pyrococcus horikoshii.

```
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000007; BAA30965.1; -.
DR   PIR; F71196; F71196.
DR   GO; GO:0009058; P:biosynthesis; IEA.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   381 AA;  42826 MW;  ED0EB0CA0CF5CAE1 CRC64;

  Query Match             13.0%;  Score 90;  DB 2;  Length 381;
  Best Local Similarity   26.8%;  Pred. No. 1.8;
  Matches   34;  Conservative   23;  Mismatches   38;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRL----SNFP--------------RLPGILAPETVLLPFCYKVF 53
              |||:   :::|:  | :    :| |               ::|||::|        :
Db         24 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 73

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                :| |::  ||: |: |  ||||  |||       | :| | : |:: :| |
Db         74 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132

Qy        112 QLTQTLG 118
              :|  |||
Db        133 KLWDTLG 139


RESULT 3
P93825
ID   P93825      PRELIMINARY;      PRT;   645 AA.
AC   P93825;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   F19P19.25 protein (Putative ethylene receptor).
GN   Name=F19P19.25; Synonyms=ERS2;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Osborne B.I., Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G.,
RA   Buehler E., Conway A.B., Conway A.R., Dewar K., Feng J., Kim C.,
RA   Kurtz D., Li Y., Shinn P., Sun H., Davis R.W., Ecker J.R.,
RA   Federspiel N.A., Theologis A.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Hua J., Sakai1 H., Nourizadeh S., Chen Q.G., Bleecker A.B.,
RA   Ecker J.R., Meyerowitz E.M.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Sakai H., Hua J., Chen Q.G., Chang C., Medrano L.J., Bleecker A.B.,
RA   Meyerowitz E.M.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC000104; AAB70445.1; -.
DR   EMBL; AF047976; AAC62209.1; -.
DR   PIR; F86174; F86174.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR   GO; GO:0007165; P:signal transduction; IEA.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR011052; Prot_amyl_inhib.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00388; HisKA; 1.
KW   Receptor.
SQ   SEQUENCE   645 AA;  72192 MW;  D37ABE3A0939D6ED CRC64;

  Query Match             11.9%;  Score 82.5;  DB 2;  Length 645;
  Best Local Similarity   23.5%;  Pred. No. 17;
  Matches   31;  Conservative   26;  Mismatches   46;  Indels   29;  Gaps    5;

Qy         25 LRHRLSNF---PRLPGILAPETV-----------------LLPFCYKVFRKKEKVKRSQ 63
              : | |: |   |  | ::   ||                 |||   |   |  :   |:
Db         99 MTHLLAGFTYGPHWPWVMTAVTVFKMLTGIVSFLTALSLVTLLPLLLKA--KVREFMLSK 156

Qy         64 KATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKC-SSLSSEAILFTLTLQLTQTLGLECC 122
              |  |     :::    |:   :| | ::  : |  :||    ||:|  ::|::||||: |
Db        157 KTRE-----LDREVGIIMKQTETSLHVRMLTTKIRTSLDRHTILYTTLVELSKTLGLKNC 211

Qy        123 LLYLSKTIHPQI 134
              :::   |  ::
Db        212 AVWIPNEIKTEM 223


RESULT 4
Q7SZX8
ID   Q7SZX8      PRELIMINARY;      PRT;   451 AA.
AC   Q7SZX8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   SI:zK13A21.9 (Novel protein similar to vertebrate microphthalmia-
DE   associated transcription factor (MITF) and zebrafish transcription
DE   factor binding to IGHM enhancer 3a (Tfe3a)).
GN   Name=opn1lw2; Synonyms=SI:zK13A21.9;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pandian R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
DR   EMBL; AL844847; CAE30419.1; -.
DR   HSSP; P22415; 1AN4.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 06:16:18 ; Search time 119.463 Seconds
(without alignments)
16226.276 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174357_19482/app_query.fasta_1.4942
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_232_@runat_07092005_174357_19482 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		694	15.9	135	3	AAY59296	Aay59296 Prostate
2		694	15.9	135	3	AAY59295	Aay59295 Prostate
3		694	15.9	135	4	AAU02174	Aau02174 Biomarker
4		694	15.9	135	4	AAU02175	Aau02175 Biomarker
c 5		313.5	7.2	129	8	ADR08959	Adr08959 Human pro
c 6		308.5	7.1	110	3	AAB43292	Aab43292 Human ORF
c 7		304.5	7.0	95	4	AAO06985	Aao06985 Human pol
c 8		302	7.0	118	8	ADQ67624	Adq67624 Novel hum
c 9		300.5	6.9	119	5	ADK37016	Adk37016 Novel hum
c 10		298.5	6.9	100	4	AAB92614	Aab92614 Human pro
	11	296.5	6.8	7036	6	AAE35264	Aae35264 Human P45
c 12		295.5	6.8	126	5	ABG65578	Abg65578 Human bre
c 13		293.5	6.8	98	5	ADG79629	Adg79629 Human sec
c 14		293.5	6.8	341	4	AAU27917	Aau27917 Human con
c 15		291.5	6.7	102	4	AAO08307	Aao08307 Human pol
	16	291	6.7	119	4	AAM83429	Aam83429 Human imm
c 17		290.5	6.7	97	8	ADM87815	Adm87815 Human EST
c 18		290.5	6.7	202	4	AAE11962	Aae11962 Human c-m
	19	289	6.6	151	5	AAU76310	Aau76310 Human fas
c 20		288.5	6.7	129	8	ADQ65903	Adq65903 Novel hum
c 21		288	6.7	164	4	AAE07113	Aae07113 Human gen
c 22		287.5	6.6	116	5	ADG79555	Adg79555 Human sec
c 23		287.5	6.6	116	5	ADG79456	Adg79456 Human sec
c 24		286.5	6.6	164	8	ADQ66320	Adq66320 Novel hum
c 25		285	6.6	107	4	AAB94965	Aab94965 Human pro
c 26		284.5	6.6	144	5	ADK37055	Adk37055 Novel hum
c 27		284.5	6.6	170	7	ADC86487	Adc86487 Human GPC
c 28		283.5	6.6	116	6	AAO26240	Aao26240 MDDT rela
c 29		283.5	6.6	119	4	AAO06149	Aao06149 Human pol
c 30		283.5	6.6	128	4	ABB11824	Abb11824 Human sec
c 31		283.5	6.6	135	7	ADE08953	Ade08953 Novel pro
c 32		283.5	6.6	135	8	ADS11754	Ads11754 Human the
c 33		283.5	6.6	152	7	ADB37629	Adb37629 Neural th
c 34		283	6.5	127	7	ADB37543	Adb37543 Neural th
c 35		282.5	6.5	99	4	AAG78806	Aag78806 Human neu
c 36		282.5	6.5	99	5	ADK36872	Adk36872 Novel hum
c 37		282.5	6.5	119	4	AAU86631	Aau86631 Novel hum
c 38		282.5	6.5	119	7	ADB59965	Adb59965 Connectiv
c 39		282.5	6.5	127	7	ADM03873	Adm03873 Human pro
c 40		282	6.5	129	4	AAB95167	Aab95167 Human pro
c 41		282	6.5	129	7	ADB37541	Adb37541 Neural th
c 42		282	6.5	136	7	ADB65293	Adb65293 Human pro
c 43		281.5	6.5	115	4	AAU32097	Aau32097 Novel hum
c 44		281.5	6.5	125	8	ADR08586	Adr08586 Human pro
c 45		281.5	6.5	260	8	ADR09618	Adr09618 Human pro

ALIGNMENTS

RESULT 1
AAY59296
ID AAY59296 standard; peptide; 135 AA.
XX
AC AAY59296;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AAZ87584.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 184-186; 191pp; English.

```
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 135 AA;

Alignment Scores:
Pred. No.:                3.14e-73       Length:          135
Score:                    694.00         Matches:         135
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              15.86%         Indels:          0
DB:                       3              Gaps:            0

US-09-974-546C-85 (1-2506) x AAY59296 (1-135)

Qy         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2
AAY59295
ID   AAY59295 standard; peptide; 135 AA.
XX
AC   AAY59295;
XX
DT   19-APR-2000  (first entry)
XX
DE   Prostate disease marker UC Band #28 amino acid sequence.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
DR   N-PSDB; AAZ87583.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Example 5; Page 182-183; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 135 AA;

Alignment Scores:
Pred. No.:                3.14e-73       Length:          135
Score:                    694.00         Matches:         135
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              15.86%         Indels:          0
DB:                       3              Gaps:            0

US-09-974-546C-85 (1-2506) x AAY59295 (1-135)

Qy         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 11:23:58 ; Search time 25.911 Seconds
(without alignments)
14439.486 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174359_19522/app_query.fasta_1.4942
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_42_@runat_07092005_174359_19522 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	771	17.6	157	4	US-09-949-016-9182	Sequence 9182, Ap
2	694	15.9	135	3	US-09-097-199-84	Sequence 84, Appl
3	694	15.9	135	3	US-09-097-199-86	Sequence 86, Appl
c 4	278.5	6.4	310	4	US-09-800-729-190	Sequence 190, App
c 5	277.5	6.4	239	4	US-09-800-729-193	Sequence 193, App
c 6	263	6.1	144	4	US-09-513-999C-6953	Sequence 6953, Ap
c 7	252.5	5.8	108	4	US-09-513-999C-4433	Sequence 4433, Ap
c 8	249.5	5.8	375	2	US-08-454-557C-121	Sequence 121, App
c 9	249.5	5.8	375	2	US-08-340-426D-121	Sequence 121, App
c 10	249.5	5.8	375	2	US-08-450-673C-121	Sequence 121, App
c 11	249.5	5.8	375	4	US-09-872-968-2	Sequence 2, Appli
c 12	248.5	5.7	96	4	US-09-513-999C-6065	Sequence 6065, Ap
13	240.5	5.5	95	4	US-09-621-976-3964	Sequence 3964, Ap
c 14	240	5.5	169	4	US-09-663-600A-208	Sequence 208, App
c 15	235.5	5.4	119	4	US-09-513-999C-7867	Sequence 7867, Ap
c 16	231	5.3	92	4	US-09-248-796A-16335	Sequence 16335, A
c 17	230	5.3	99	4	US-09-288-143-168	Sequence 168, App
c 18	229.5	5.3	112	4	US-09-513-999C-7870	Sequence 7870, Ap
c 19	229.5	5.3	118	4	US-09-663-600A-114	Sequence 114, App
c 20	224	5.2	1079	3	US-09-058-489-22	Sequence 22, Appl
c 21	223	5.2	397	5	PCT-US95-17111A-121	Sequence 121, App
c 22	221.5	5.1	132	4	US-09-636-215-573	Sequence 573, App
c 23	221.5	5.1	132	4	US-09-685-166A-573	Sequence 573, App
c 24	221.5	5.1	132	4	US-09-679-426-573	Sequence 573, App
c 25	221.5	5.1	132	4	US-09-759-143-573	Sequence 573, App
c 26	221.5	5.1	132	4	US-09-651-236-573	Sequence 573, App
c 27	221.5	5.1	135	4	US-09-685-166A-884	Sequence 884, App
c 28	221.5	5.1	135	4	US-09-679-426-884	Sequence 884, App
c 29	221.5	5.1	135	4	US-09-759-143-884	Sequence 884, App
c 30	220	5.1	91	4	US-09-621-976-5929	Sequence 5929, Ap
31	220	5.0	107	4	US-09-513-999C-7879	Sequence 7879, Ap
c 32	218	5.0	121	4	US-09-513-999C-7874	Sequence 7874, Ap
33	217.5	5.0	84	4	US-09-621-976-4377	Sequence 4377, Ap
c 34	216	5.0	92	4	US-09-621-976-5895	Sequence 5895, Ap
c 35	216	5.0	92	4	US-09-621-976-5901	Sequence 5901, Ap
c 36	216	5.0	92	4	US-09-621-976-5909	Sequence 5909, Ap
c 37	216	5.0	92	4	US-09-621-976-5914	Sequence 5914, Ap
c 38	212.5	4.9	776	4	US-10-020-079-24	Sequence 24, Appl
c 39	212.5	4.9	863	4	US-10-020-079-32	Sequence 32, Appl
c 40	212.5	4.9	876	4	US-10-020-079-30	Sequence 30, Appl
c 41	212.5	4.9	895	4	US-10-020-079-18	Sequence 18, Appl
c 42	212.5	4.9	976	4	US-10-020-079-28	Sequence 28, Appl
c 43	212.5	4.9	982	4	US-10-020-079-26	Sequence 26, Appl
c 44	210.5	4.9	83	4	US-09-621-976-5396	Sequence 5396, Ap
c 45	208.5	4.8	789	4	US-10-020-079-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9182
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9182

Alignment Scores:
Pred. No.: 1.03e-84 Length: 157
Score: 771.00 Matches: 155
Percent Similarity: 97.48% Conservative: 0
Best Local Similarity: 97.48% Mismatches: 2
Query Match: 17.62% Indels: 2
DB: 4 Gaps: 0

US-09-974-546C-85 (1-2506) x US-09-949-016-9182 (1-157)

Qy 31 TGTATAATAATTTACAAAATTATTCTTCTATTGCTACAGAGCTACAATTCAATTTACAGT 90

```
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 CysIleIleIleTyrLysIleIleLeuLeu-CysTyrArgAlaThrIleGlnPheThrVa 20

Qy         91 AGGCCACCATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTC 150
                |       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 l-ValAsnMetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleH 40

Qy        151 ACATTTTACAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCA 210
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         40 isIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyI 60

Qy        211 TTCTAGCTCCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAA 270
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 leLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluL 80

Qy        271 AAGTAAAAAGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACC 330
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 ysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisH 100

Qy        331 ATGCAATTCTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTT 390
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 isAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysS 120

Qy        391 CCTCATTATCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAG 450
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 erSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuG 140

Qy        451 GTCTGGAATGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 lyLeuGluCysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 157

RESULT 2
US-09-097-199-84
; Sequence 84, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  84:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-097-199-84

Alignment Scores:
Pred. No.:              2.33e-75        Length:         135
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            15.86%          Indels:         0
DB:                     3               Gaps:           0

US-09-974-546C-85 (1-2506) x US-09-097-199-84 (1-135)

Qy         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
                ||||||||||||||||||||||||||||||||||||||||||||
Db        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 3
US-09-097-199-86
; Sequence 86, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 12:21:48 ; Search time 141.556 Seconds
(without alignments)
13965.459 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 3554922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174400_19612/app_query.fasta_1.4942
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09974546_@CGN_1_1_229_@runat_07092005_174400_19612
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	15.9	135	10	US-09-974-546-84	Sequence 84, Appl
2	694	15.9	135	10	US-09-974-546-86	Sequence 86, Appl
c 3	301.5	7.0	109	15	US-10-121-016-48	Sequence 48, Appl
c 4	295.5	6.8	126	14	US-10-082-830-207	Sequence 207, App
c 5	293.5	6.8	341	18	US-10-220-335-586	Sequence 586, App
c 6	290.5	6.7	97	15	US-10-112-944-908	Sequence 908, App
c 7	284.5	6.6	170	14	US-10-017-161-1108	Sequence 1108, Ap
c 8	284.5	6.6	170	15	US-10-292-798-940	Sequence 940, App
c 9	283.5	6.6	128	15	US-10-276-774-2194	Sequence 2194, Ap
c 10	283.5	6.6	152	14	US-10-198-070-112	Sequence 112, App
c 11	283	6.5	127	14	US-10-198-070-25	Sequence 25, Appl
c 12	282.5	6.5	119	9	US-09-764-847-696	Sequence 696, App
c 13	282.5	6.5	119	14	US-10-092-154-696	Sequence 696, App
c 14	282.5	6.5	127	15	US-10-108-260A-2558	Sequence 2558, Ap
c 15	282	6.5	129	14	US-10-198-070-23	Sequence 23, Appl
c 16	282	6.5	136	15	US-10-104-047-3447	Sequence 3447, Ap
c 17	281	6.5	120	15	US-10-108-260A-4121	Sequence 4121, Ap
c 18	280.5	6.5	144	14	US-10-011-585A-145	Sequence 145, App
c 19	279.5	6.5	134	15	US-10-416-314-45	Sequence 45, Appl
c 20	279	6.4	133	15	US-10-276-774-1788	Sequence 1788, Ap
21	278.5	6.4	136	15	US-10-094-749-1763	Sequence 1763, Ap
c 22	278.5	6.4	310	9	US-09-800-729-190	Sequence 190, App
c 23	277.5	6.4	127	11	US-09-833-245-1427	Sequence 1427, Ap
c 24	277.5	6.4	239	9	US-09-800-729-193	Sequence 193, App
c 25	277.5	6.4	239	14	US-10-198-070-113	Sequence 113, App
c 26	277.5	6.4	239	18	US-10-756-149-5485	Sequence 5485, Ap
c 27	275.5	6.4	109	9	US-09-989-919-105	Sequence 105, App
c 28	275.5	6.4	156	15	US-10-104-047-1984	Sequence 1984, Ap
c 29	275.5	6.4	170	10	US-09-986-480-282	Sequence 282, App
c 30	275.5	6.4	170	17	US-10-863-332-282	Sequence 282, App
c 31	275.5	6.4	198	10	US-09-986-480-433	Sequence 433, App
c 32	275.5	6.4	198	17	US-10-863-332-433	Sequence 433, App
c 33	274.5	6.3	100	13	US-10-016-157A-187	Sequence 187, App
c 34	274.5	6.3	134	15	US-10-276-774-2463	Sequence 2463, Ap
c 35	274	6.3	167	14	US-10-017-161-1154	Sequence 1154, Ap
c 36	274	6.3	167	15	US-10-292-798-974	Sequence 974, App
c 37	273.5	6.3	107	9	US-09-989-920-233	Sequence 233, App
c 38	273.5	6.3	241	18	US-10-486-977-2	Sequence 2, Appli
c 39	272	6.3	122	11	US-09-833-245-1659	Sequence 1659, Ap
c 40	271	6.3	171	18	US-10-496-905-352	Sequence 352, App
c 41	270	6.2	190	14	US-10-017-161-1734	Sequence 1734, Ap
c 42	270	6.2	190	15	US-10-292-798-1390	Sequence 1390, Ap
c 43	269.5	6.2	116	15	US-10-108-260A-2525	Sequence 2525, Ap
c 44	269.5	6.2	123	15	US-10-104-047-2112	Sequence 2112, Ap
c 45	269.5	6.2	125	15	US-10-108-260A-4066	Sequence 4066, Ap

ALIGNMENTS

RESULT 1
US-09-974-546-84
; Sequence 84, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:

```
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/974,546
;              FILING DATE: 10-Oct-2001
;              CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 09/097,199
;              FILING DATE: 1998-06-12
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Nakashima, Richard A.
;              REGISTRATION NUMBER: P-42,023
;              REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (512) 418-3000
;              TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 84:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 135 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

Alignment Scores:
Pred. No.:                6.08e-71        Length:         135
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              15.86%          Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-85 (1-2506) x US-09-974-546-84 (1-135)

Qy          99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy         159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy         219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy         279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy         339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy         399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy         459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
               |||||||||||||||||||||||||||||||||||||||||||
Db         121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2
US-09-974-546-86
; Sequence 86, Application US/09974546
; Publication No. US20030050470A1
;   GENERAL INFORMATION:
;        APPLICANT: An, Gang
;                   O'Hara, S. Mark
;                   Ralph, David
;                   Veltri, Robert
;        TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                            PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/974,546
;             FILING DATE: 10-Oct-2001
;             CLASSIFICATION: Unknown
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/097,199
;             FILING DATE: 1998-06-12
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 86:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 135 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Alignment Scores:
Pred. No.:                6.08e-71        Length:         135
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              15.86%          Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-85 (1-2506) x US-09-974-546-86 (1-135)

Qy          99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Qy         159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy         219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy         279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy         339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy         399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 09:43:33 ; Search time 30.2749 Seconds
(without alignments)
15928.642 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174358_19508/app_query.fasta_1.4942
-DB=PIR_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_79_@runat_07092005_174358_19508 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	239.5	5.5	613	4	C40201	artifact-warning s
	2	231.5	5.3	613	4	C40201	artifact-warning s
c	3	228	5.3	597	4	E40201	artifact-warning s
	4	224	5.1	597	4	E40201	artifact-warning s
c	5	204	4.7	627	4	A40201	artifact-warning s
	6	194.5	4.4	673	4	F40201	artifact-warning s
c	7	190.5	4.4	673	4	F40201	artifact-warning s
	8	187.5	4.3	627	4	A40201	artifact-warning s
	9	173.5	4.0	196	2	I38022	hypothetical prote
	10	161	3.7	579	4	D40201	artifact-warning s
	11	149	3.4	418	2	S41044	chromosomal protei
c	12	145	3.4	301	4	B40201	artifact-warning s
c	13	142	3.3	579	4	D40201	artifact-warning s
c	14	141.5	3.3	100	2	A46010	X-linked retinopat
	15	133.5	3.1	114	2	JC5238	galactosylceramide
	16	132.5	3.0	124	2	A47582	B-cell growth fact
c	17	129	3.0	499	2	S65657	alpha-1C-adrenergi
	18	120.5	2.8	301	4	B40201	artifact-warning s
c	19	120	2.8	342	2	PC4211	hepatocellular car
c	20	120	2.8	841	1	I78885	serine/threonine-s
	21	103.5	2.4	381	2	F71196	probable hexosyltr
	22	103	2.4	1613	2	D90129	hypothetical prote
c	23	100.5	2.3	369	2	A53959	thromboxane A-2 re
c	24	100.5	2.3	619	2	A60646	transforming prote
	25	98	2.2	1715	2	T30559	resistance protein
c	26	93.5	2.2	79	2	A56194	thromboxane A-2 re
c	27	93.5	2.2	407	2	T02670	probable thromboxa
	28	92.5	2.1	263	2	A82069	hypothetical prote
	29	92	2.1	461	1	GQRTT1	tumor necrosis fac
	30	92	2.1	536	2	T37840	probable vacuolar
	31	90	2.1	2539	2	B71619	hypothetical prote
	32	89.5	2.0	296	2	S01441	class II histocomp
	33	89.5	2.0	399	2	D86322	hypothetical prote
c	34	88.5	2.0	252	2	PC4259	ferritin associate
	35	88.5	2.0	795	2	S30874	hypothetical prote
	36	87.5	2.0	863	2	C88546	protein R107.4 [im
	37	87	2.0	379	2	C75006	hypothetical prote
	38	87	2.0	493	2	S46826	hypothetical prote
	39	86.5	2.0	451	2	D70045	two-component sens
	40	86.5	2.0	459	2	S06607	23S rRNA intron 2
	41	86	2.0	699	2	S51454	EST1 protein - yea
	42	85.5	2.0	523	2	D64555	conserved hypothet
	43	85.5	2.0	679	2	B96599	protein F20N2.12 [
	44	85.5	2.0	1386	1	RNLVC2	DNA-directed RNA p
	45	85.5	2.0	3119	2	T18414	protein g377 - mal

ALIGNMENTS

RESULT 1
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi
.

Alignment Scores:
Pred. No.: 3.88e-17 Length: 613
Score: 239.50 Matches: 61
Percent Similarity: 73.63% Conservative: 6
Best Local Similarity: 67.03% Mismatches: 17
Query Match: 5.54% Indels: 8
DB: 4 Gaps: 2

US-09-974-546C-85 (1-2506) x C40201 (1-613)

```
Qy       2483 GTTGCCCAGGCTGGAGTGCAATAGTGCAATCTCAGCTCACCGCAACAACCTCCACCTTCC 2424
                  |||   |||||||||||||||      :::|||   |||   |||   |||||||||
Db        322 ValThrGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProPro--- 340

Qy       2423 GGGTTCAAGCAATTCTCCGC-CTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGTGCCA 2365
                  ||||||      ||||||   ||||||||||||   :::||||||||||||        ||
```

```
Db        341 GlyPheMetProPheSerCysLeuSerLeuLeuArgThrTrpAspTyrArgArgPro-Hi 360

Qy       2364 CCACGCCCGGCTAATTTTTTGTGTTTTTAGTAGAGATGGGGGTTCCACCATGCTGGTCAG 2305
             |||||||:::|||||||||||||:::|||||||||::: ||||||   |||||||||   |
Db        360 sHisAlaGlnLeuIlePheCysIlePheSerArgAsn-GlyValLeuProCysTrpProG 380

Qy       2304 GCTGGTCTCGAACTCCCGACCTCAGGTGATCCACCCA------CCTCCCAAAGTGCTGGG 2251
             ||||||||||||||||   |||  :::|||||||||      ||||||||||||||||
Db        380 lyTrpSerArgThrProGlyLeu--MetIleHisProProLeuProProLysValLeuGl 399

Qy       2250 ATTACAGGCGTGAGCCACCGTGCCT 2226
             ||||||||||   ||||||   |||
Db        399 yLeuGlnAla***AlaThrAlaPro 407

RESULT 2
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              2.91e-16        Length:         613
Score:                  231.50          Matches:        53
Percent Similarity:     63.33%          Conservative:   4
Best Local Similarity:  58.89%          Mismatches:     27
Query Match:            5.29%           Indels:         6
DB:                     4               Gaps:           2

US-09-974-546C-85 (1-2506) x C40201 (1-613)

Qy       2227 GGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA------GGTGGGTGGATCACC 2280
             |||   |||||||||||||||||||||||||||||||      |||||||||||  :
Db        208 GlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyLysGlyGlyTrpIle--M 227

Qy       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
             ::|||   ||||||||||||||||   |||||||||   ||||||   |||||||||
Db        227 etArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPhe-LeuLeuLys 246

Qy       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
                |||||||||||   ||||||                                :::
Db        247 IleGlnLysIleSerTrpAlaTrpCysGlyArgLeu***SerHisValArgArgArgLeu 266

Qy       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
                ||||||   :::||||||   |||||||||   |||   |||   |||
Db        267 ArgGlnGluAsnGlyIleAsnPro---GlyGlyGlyAlaCysSerGluProArgSerArg 285

Qy       2461 TATTGCACTCCAGCCTGGGCAACA 2484
             :::|||||||||||||||   |||
Db        286 HisCysThrProAlaTrpValThr 293

RESULT 3
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi
.

Alignment Scores:
Pred. No.:              6.99e-16        Length:         597
Score:                  228.00          Matches:        65
Percent Similarity:     72.45%          Conservative:   6
Best Local Similarity:  66.33%          Mismatches:     16
Query Match:            5.27%           Indels:         13
DB:                     4               Gaps:           4

US-09-974-546C-85 (1-2506) x E40201 (1-597)

Qy       2495 AGTTTTACTCCTGTTGCCCAGGCTGGAGTGCAATAGTGCAATCTCAGCTCACC----GCA 2440
             ||||||      |||||||||||||||||:::      ||||||::::::      |||
Db        507 SerPheAlaLeuValAlaGlnAlaGlyValArgTrpHisAsnLeuThrAlaAsnPheAla 526

Qy       2439 ACAACCTCCACCTTCCGGGTTCAAGCAATTCTCCGC---CTCAGCCTCCTGAGTAGCTGG 2383
             :::                |||||||||||||||      |||||||||   |||||||||
Db        527 SerTrp------------ValGlnAlaIleLeuSerCysLeuSerLeuProSerSerTrp 542

Qy       2382 GATTACAGGCATGTGCCACCACGCCCGGCTAATTTTTTGTGTTTTTAGTAGAGATGGGGG 2323
             ||||||||||||   ||||||||||||||||||  |||:::|||||||||||||||| |||
Db        543 AspTyrArgHisAlaProProArgProAlaAsn--PheIlePheLeuValGluMet-Gly 561

Qy       2322 TTCCACCATGCTGGTCAGGCTGGTCTCGAACTCCCGACCTCAGGTGATCCACCCAC---- 2267
             |||   |||   |||||||||||||||:::|||||||||||||||||||||||||
Db        562 PheLeuHisValGlyGlnAlaGlyLeuLysLeuProThrSerGlyAspProProArgLeu 581

Qy       2266 ---CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGTGCCTGGCC 2222
                |||||||||   ||||||||||||   |||   |||   ||||||
Db        582 GluLeuProLysArgTrpAspTyrArgHisGluLeuProHisLeuAla 597

RESULT 4
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi
.

Alignment Scores:
Pred. No.:              1.92e-15        Length:         597
Score:                  224.00          Matches:        60
Percent Similarity:     68.04%          Conservative:   6
Best Local Similarity:  61.86%          Mismatches:     23
Query Match:            5.12%           Indels:         9
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 06:29:13 ; Search time 149.193 Seconds
(without alignments)
17202.865 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174358_19494/app_query.fasta_1.4942
-DB=UniProt_03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_302_@runat_07092005_174358_19494 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	694	15.9	135	2	Q9GZY1	Q9gzy1 homo sapien
	2	333.5	7.6	593	1	ALU6_HUMAN	P39193 homo sapien
c	3	327	7.6	593	1	ALU6_HUMAN	P39193 homo sapien
	4	316.5	7.2	593	1	ALU7_HUMAN	P39194 homo sapien
c	5	313.5	7.2	129	2	Q6ZSV6	Q6zsv6 homo sapien
c	6	308	7.1	593	1	ALU7_HUMAN	P39194 homo sapien
c	7	297.5	6.9	140	2	Q6ZP99	Q6zp99 homo sapien
c	8	288.5	6.7	129	2	Q6ZUW8	Q6zuw8 homo sapien
c	9	286.5	6.6	164	2	Q6ZUK0	Q6zuk0 homo sapien
c	10	283.5	6.6	152	2	Q9NX85	Q9nx85 homo sapien
c	11	283	6.5	127	2	Q9H743	Q9h743 homo sapien
c	12	282	6.5	129	2	Q9H9H0	Q9h9h0 homo sapien
c	13	281.5	6.5	125	2	Q6ZTE1	Q6zte1 homo sapien
c	14	281.5	6.5	260	2	Q6ZQR8	Q6zqr8 homo sapien
c	15	281	6.5	120	2	Q8N8A3	Q8n8a3 homo sapien
c	16	280	6.5	132	2	Q6ZNU7	Q6znu7 homo sapien
c	17	278.5	6.4	134	2	Q6ZSG8	Q6zsg8 homo sapien
	18	278.5	6.4	136	2	Q96NR6	Q96nr6 homo sapien
c	19	278.5	6.4	146	2	Q6ZTX8	Q6ztx8 homo sapien
c	20	277.5	6.4	239	2	Q9NX17	Q9nx17 homo sapien
	21	277	6.3	591	1	ALU8_HUMAN	P39195 homo sapien
c	22	276	6.4	132	2	Q6ZS28	Q6zs28 homo sapien
c	23	276	6.4	591	1	ALU8_HUMAN	P39195 homo sapien
	24	275.5	6.3	131	2	Q6ZNZ7	Q6znz7 homo sapien
c	25	275.5	6.4	156	2	Q8NBH4	Q8nbh4 homo sapien
c	26	274.5	6.3	142	2	Q6ZWD5	Q6zwd5 homo sapien
c	27	271.5	6.3	239	2	Q6ZWA9	Q6zwa9 homo sapien
c	28	269.5	6.2	125	2	Q8N8C2	Q8n8c2 homo sapien
c	29	269	6.2	148	2	Q6ZUC5	Q6zuc5 homo sapien
c	30	268	6.2	118	2	Q9H387	Q9h387 homo sapien
c	31	267	6.2	202	2	Q6ZUA4	Q6zua4 homo sapien
c	32	266.5	6.2	130	2	Q6ZSV2	Q6zsv2 homo sapien
c	33	266	6.1	123	2	Q9HAD8	Q9had8 homo sapien
c	34	264	6.1	180	2	Q8N7M7	Q8n7m7 homo sapien
c	35	263.5	6.1	122	2	Q6ZNZ6	Q6znz6 homo sapien
c	36	263.5	6.1	128	2	Q6ZPB2	Q6zpb2 homo sapien
	37	260.5	6.0	123	2	Q9P1N7	Q9p1n7 homo sapien
c	38	260.5	6.0	146	2	Q6ZRX6	Q6zrx6 homo sapien
c	39	260.5	6.0	152	2	Q6ZUG4	Q6zug4 homo sapien
c	40	260	6.0	124	2	Q6ZUN5	Q6zun5 homo sapien
c	41	260	6.0	585	1	ALU5_HUMAN	P39192 homo sapien
c	42	258.5	6.0	142	2	Q6ZRQ0	Q6zrq0 homo sapien
c	43	258.5	6.0	179	2	Q8N1K7	Q8n1k7 homo sapien
c	44	257.5	6.0	167	2	Q6ZSN3	Q6zsn3 homo sapien
c	45	257.5	6.0	174	2	Q8N2A0	Q8n2a0 homo sapien

ALIGNMENTS

RESULT 1
Q9GZY1
ID Q9GZY1 PRELIMINARY; PRT; 135 AA.
AC Q9GZY1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RA Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1; -.
DR EMBL; AF189269; AAG17117.1; -.
DR Genew; HGNC:21079; PBOV1.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B1983705D0 CRC64;

Alignment Scores:
Pred. No.: 4.25e-63 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.86% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-85 (1-2506) x Q9GZY1 (1-135)

Qy 99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20

```
Qy        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Qy        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Qy        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80

Qy        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Qy        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Qy        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
```

```
RESULT 2
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   DOMAIN       1     97       Frame-1.
FT   DOMAIN     101    196       Frame-2.
FT   DOMAIN     200    295       Frame-3.
FT   DOMAIN     299    395       Frame-4.
FT   DOMAIN     399    494       Frame-5.
FT   DOMAIN     498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64603 MW;  136EF344AACD12A2 CRC64;
```

```
Alignment Scores:
Pred. No.:              2.9e-25        Length:          593
Score:                  333.50         Matches:         76
Percent Similarity:     84.38%         Conservative:    5
Best Local Similarity:  79.17%         Mismatches:      10
Query Match:            7.62%          Indels:          6
DB:                     1              Gaps:            2

US-09-974-546C-85 (1-2506) x ALU6_HUMAN (1-593)

Qy       2222 GGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGG------GAGGTGGGTGGA 2275
              |||:::|||||||||||||||||||||||||||||||||||      |||   ||||||
Db          1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20

Qy       2276 TCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTAC 2335
              |||||||||||||||||||||||||||||   |||   |||   ||| |||:::|||||
Db         21 SerProGluValGlySerSerArgProAla***ProThrTrpArgAsn-ProValSerTh 40

Qy       2336 TAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGG 2395
              |||||||||| ||||||||||||||||||   |||||||||||||||||||:::|
Db         40 rLysAsnThr-LysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgG 60

Qy       2396 AGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGA 2454
              ||||||||   ||||||   ||||||||||||||   |||   ||||||   |||||||
Db         60 luAlaGluAlaGlyGluSerLeuGluProGlyArg---ArgArgLeuArg***AlaGluI 79

Qy       2455 TTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCC 2498
              |||||   |||||||||||||||||:::|||:::||||||
Db         79 leAlaProLeuHisSerSerLeuGlyAsnLysSerGluThrPro 93
```

```
RESULT 3
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:24:09 ; Search time 7220.71 Seconds
(without alignments)
16816.746 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactccccccaccc 2506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2505.6	100.0	2512	9	AF189269	AF189269 Homo sapi
	2	2494	99.5	2505	6	AR146836	AR146836 Sequence
	3	2424.2	96.7	129227	9	HS171N11	AL031433 Human DNA
	4	2424.2	96.7	186698	2	AC036236	AC036236 Homo sapi
	5	2417	96.4	2496	9	BC069109	BC069109 Homo sapi
	6	1961.6	78.3	2103	9	AF189270	AF189270 Homo sapi
	7	1950	77.8	2087	6	AR146835	AR146835 Sequence
	8	626.2	25.0	757	6	AR146757	AR146757 Sequence
c	9	201.6	8.0	137227	9	AC092933	AC092933 Homo sapi
	10	201.6	8.0	166895	2	AC023415	AC023415 Homo sapi
	11	201.6	8.0	170961	2	AC112202	AC112202 Homo sapi
c	12	201.6	8.0	174054	2	AC131148	AC131148 Homo sapi
	13	198.8	7.9	200237	9	AF168787	AF168787 Homo sapi
c	14	198.6	7.9	129625	9	AC009301	AC009301 Homo sapi
	15	198.6	7.9	318586	2	AC107422	AC107422 Homo sapi
	16	195.6	7.8	73499	9	AC132942	AC132942 Homo sapi
c	17	195.6	7.8	209885	2	AC011189	AC011189 Homo sapi
	18	194.4	7.8	120551	9	AL391215	AL391215 Human DNA
c	19	194.2	7.7	156886	2	AC023445	AC023445 Homo sapi
c	20	194.2	7.7	174551	9	AC010598	AC010598 Homo sapi
	21	193	7.7	56884	2	AC104588	AC104588 Homo sapi
c	22	192.8	7.7	164107	9	AC104982	AC104982 Homo sapi
c	23	192.4	7.7	150636	2	AC027529	AC027529 Homo sapi
	24	192.4	7.7	192203	9	AC034102	AC034102 Homo sapi
	25	192.2	7.7	146957	9	AC084264	AC084264 Homo sapi
c	26	192.2	7.7	180346	9	AC092615	AC092615 Homo sapi
	27	191.8	7.7	131637	9	CNS01DUC	AL133239 Human chr
	28	191.8	7.7	164514	9	CNS01RGV	AL160211 Human chr
	29	191.2	7.6	110000	2	BX324168_0	BX324168 Homo sapi
	30	191.2	7.6	112884	9	AL136966	AL136966 Human DNA
c	31	191.2	7.6	162377	9	AL683870	AL683870 Human DNA
c	32	191	7.6	187490	2	AC048378	AC048378 Homo sapi
	33	190.6	7.6	109588	9	AC010461	AC010461 Homo sapi
	34	190.6	7.6	161920	9	AC025169	AC025169 Homo sapi
	35	190.4	7.6	148984	9	HS105D16	AL031311 Human DNA
c	36	190.2	7.6	182945	2	AC069004	AC069004 Homo sapi
	37	190.2	7.6	210643	9	AC137723	AC137723 Homo sapi
	38	190	7.6	128379	9	HSJ773A18	AL049557 Human DNA
c	39	190	7.6	163521	2	AC108040	AC108040 Homo sapi
	40	190	7.6	196773	9	AC114480	AC114480 Homo sapi
	41	189.8	7.6	201939	9	AC044792	AC044792 Homo sapi
	42	189.6	7.6	170502	9	AC093573	AC093573 Pan trogl
c	43	189.4	7.6	147859	2	AC011079	AC011079 Homo sapi
	44	189.4	7.6	157289	2	AC021447	AC021447 Homo sapi
	45	189.4	7.6	178305	2	AC139475	AC139475 Homo sapi

ALIGNMENTS

RESULT 1
AF189269
LOCUS AF189269 2512 bp mRNA linear PRI 22-MAR-2001
DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION AF189269
VERSION AF189269.1 GI:10441601
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2512)
AUTHORS An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L., Wright,G.L. Jr. and Veltri,R.W.
TITLE Cloning and characterization of UROC28, a novel gene overexpressed in prostate, breast, and bladder cancers
JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PUBMED 11156405
REFERENCE 2 (bases 1 to 2512)
AUTHORS An,G. and Veltri,R.W.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research Parkway, Oklahoma City, OK 73104, USA
FEATURES Location/Qualifiers
source 1. .2512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
gene 1. .2512
/gene="UC28"
CDS 99. .506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17117.1"
/db_xref="GI:10441602"
/translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS

```
                 EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     polyA_signal 2487. .2492
                 /gene="UC28"
ORIGIN

  Query Match           100.0%;  Score 2505.6;  DB 9;  Length 2512;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2506;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980

Qy       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040

Qy       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100
```

```
Db       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100

Qy       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160

Qy       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280

Qy       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340

Qy       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400

Qy       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460

Qy       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
             ||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506


RESULT 2
AR146836
LOCUS       AR146836                2505 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 85 from patent US 6218529.
ACCESSION   AR146836
VERSION     AR146836.1  GI:15110025
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2505)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 85 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .2505
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             99.5%;  Score 2494;  DB 6;  Length 2505;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2505;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
                |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1979

Qy       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1980 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2039

Qy       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2040 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2099

Qy       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2100 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2159

Qy       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2160 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2219

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2220 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2279

Qy       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2280 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2339

Qy       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2340 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2399

Qy       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2400 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2459

Qy       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
                ||||||||||||||||||||||||||||||||||||||||||||||
Db       2460 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2505


RESULT 3
HS171N11
LOCUS       HS171N11              129227 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
            complete sequence.
ACCESSION   AL031433
VERSION     AL031433.4  GI:4826442
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129227)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 13, 1999 this sequence version replaced gi:4582115.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-171N11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .129227
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q23"
                     /clone="RP1-171N11"
                     /clone_lib="RPCI-1"
ORIGIN

  Query Match             96.7%;  Score 2424.2;  DB 9;  Length 129227;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 2494;  Conservative    1;  Mismatches    4;  Indels    8;  Gaps    6;
```

```
Qy        1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
Db    29796 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 29855

Qy       61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
Db    29856 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 29913

Qy      121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
Db    29914 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 29973

Qy      181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
Db    29974 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 30033

Qy      241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
Db    30034 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 30093

Qy      301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
Db    30094 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 30153

Qy      361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
Db    30154 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 30213

Qy      421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
Db    30214 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 30273

Qy      481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
Db    30274 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 30333

Qy      541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
Db    30334 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 30393

Qy      601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
Db    30394 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 30453

Qy      661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
Db    30454 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 30513

Qy      721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
Db    30514 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 30573

Qy      781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
Db    30574 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 30633

Qy      841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db    30634 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 30693

Qy      901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db    30694 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 30753

Qy      961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db    30754 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 30813

Qy     1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db    30814 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 30873

Qy     1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db    30874 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 30933

Qy     1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
Db    30934 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 30993

Qy     1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
Db    30994 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 31053

Qy     1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
Db    31054 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 31113

Qy     1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
Db    31114 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 31173

Qy     1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
Db    31174 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 31233

Qy     1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
Db    31234 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 31293

Qy     1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
Db    31294 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 31353

Qy     1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
Db    31354 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 31413

Qy     1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
Db    31414 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 31473

Qy     1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
Db    31474 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 31533

Qy     1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
Db    31534 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 31593

Qy     1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
Db    31594 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 31653

Qy     1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
Db    31654 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 31713

Qy     1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 1979
Db    31714 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 31773

Qy     1980 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 2039
Db    31774 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 31833

Qy     2040 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTC 2099
Db    31834 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTT-GTAATACCAGAACTC 31892

Qy     2100 TAATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 2159
Db    31893 TAATTGGA-CACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 31951

Qy     2160 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACC 2219
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31952 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACC 32011

Qy       2220 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 2279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32012 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 32071

Qy       2280 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 2339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32072 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 32131

Qy       2340 AACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 2399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32132 AACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 32191

Qy       2400 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCA 2459
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db      32192 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAG---GTTGCGGTGAGCTGAGATTGCA 32248

Qy       2460 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
              ||||||||||||||||||||||||||||||||||||||||||||||
Db      32249 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 32295


RESULT 4
AC036236
LOCUS       AC036236               186698 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-133O15, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC036236
VERSION     AC036236.2  GI:9958166
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:7523932.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0133O15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183373 bases at least Q40
            Consensus quality: 184432 bases at least Q30
            Consensus quality: 185064 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 186098; sum-of-contigs
            Quality coverage: 5.84 in Q20 bases; agarose-fp
            Quality coverage: 5.57 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4769: contig of 4769 bp in length
            *     4770     4869: gap of unknown length
            *     4870    11657: contig of 6788 bp in length
            *    11658    11757: gap of unknown length
            *    11758    28189: contig of 16432 bp in length
            *    28190    28289: gap of unknown length
            *    28290    53341: contig of 25052 bp in length
            *    53342    53441: gap of unknown length
            *    53442   102327: contig of 48886 bp in length
            *   102328   102427: gap of unknown length
            *   102428   182781: contig of 80354 bp in length
            *   182782   182881: gap of unknown length
            *   182882   186698: contig of 3817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186698
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-133O15"
     misc_feature    1. .4769
                     /note="assembly_name:Contig10"
     misc_feature    4870. .11657
                     /note="assembly_name:Contig11"
     misc_feature    11758. .28189
                     /note="assembly_name:Contig12"
     misc_feature    28290. .53341
                     /note="assembly_name:Contig13"
     misc_feature    53442. .102327
                     /note="assembly_name:Contig14"
     misc_feature    102428. .182781
                     /note="assembly_name:Contig15
                     clone_end:SP6
                     vector_side:right"
     misc_feature    182882. .186698
                     /note="assembly_name:Contig9"
ORIGIN

  Query Match            96.7%;  Score 2424.2;  DB 2;  Length 186698;
  Best Local Similarity  99.5%;  Pred. No. 0;
  Matches 2494;  Conservative    1;  Mismatches    4;  Indels    8;  Gaps    6;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159156 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 159215

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              |||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
Db     159216 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 159273

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159274 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 159333

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159334 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 159393

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159394 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 159453

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159454 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 159513

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159514 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 159573
```

```
Qy       421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
Db    159574 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 159633

Qy       481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
Db    159634 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 159693

Qy       541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
Db    159694 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 159753

Qy       601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
Db    159754 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 159813

Qy       661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
Db    159814 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 159873

Qy       721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
Db    159874 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 159933

Qy       781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
Db    159934 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 159993

Qy       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db    159994 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 160053

Qy       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db    160054 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 160113

Qy       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db    160114 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 160173

Qy      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db    160174 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 160233

Qy      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db    160234 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 160293

Qy      1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
Db    160294 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 160353

Qy      1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
Db    160354 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 160413

Qy      1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
Db    160414 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 160473

Qy      1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
Db    160474 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 160533

Qy      1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
Db    160534 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 160593

Qy      1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
Db    160594 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 160653

Qy      1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
Db    160654 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 160713

Qy      1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
Db    160714 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 160773

Qy      1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
Db    160774 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 160833

Qy      1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
Db    160834 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 160893

Qy      1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
Db    160894 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 160953

Qy      1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
Db    160954 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 161013

Qy      1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
Db    161014 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 161073

Qy      1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 1979
Db    161074 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 161133

Qy      1980 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 2039
Db    161134 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 161193

Qy      2040 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTC 2099
Db    161194 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTT-GTAATACCAGAACTC 161252

Qy      2100 TAATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 2159
Db    161253 TAATTGGA-CACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 161311

Qy      2160 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACC 2219
Db    161312 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACC 161371

Qy      2220 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 2279
Db    161372 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 161431

Qy      2280 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 2339
Db    161432 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 161491

Qy      2340 AACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 2399
Db    161492 AACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 161551

Qy      2400 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCA 2459
Db    161552 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAG---GTTGCGGTGAGCTGAGATTGCA 161608

Qy      2460 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
Db    161609 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 161655
```

RESULT 5
BC069109
LOCUS BC069109 2496 bp mRNA linear PRI 23-JUN-2004
DEFINITION Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cDNA clone IMAGE:7216926).

```
ACCESSION   BC069109
VERSION     BC069109.1  GI:46575743
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2496)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2496)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Anup Madan, University of Iowa
            cDNA Library Preparation: Anup Madan, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Neurogenomics Research Lab,
            200 B EMRB, University of Iowa, Iowa City, IA-52242
            anup-madan@uiowa.edu
            Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0.
FEATURES             Location/Qualifiers
     source          1. .2496
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7216926"
                     /tissue_type="Synthetic constructs"
                     /clone_lib="NIH_MGC_242"
                     /lab_host="XL10"
                     /note="Vector: pUC19"
ORIGIN

  Query Match             96.4%;  Score 2417;  DB 9;  Length 2496;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 2488;  Conservative    1;  Mismatches    6;  Indels    8;  Gaps    6;

Qy          2 ACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTAT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA- 59

Qy         62 TGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACCA 121
              |||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
Db         60 TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACCA 118

Qy        122 GAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAG 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 GAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAG 178

Qy        182 ATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATT 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 ATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATT 238

Qy        242 CTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 CTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTT 298

Qy        302 CATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTT 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTT 358

Qy        362 GACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCAC 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCAC 418

Qy        422 ATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 ATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 478

Qy        482 AACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAA 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAA 538

Qy        542 TAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 TAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACT 598

Qy        602 GAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTCA 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 GAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTCA 658

Qy        662 ATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCTA 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 ATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCTA 718

Qy        722 TTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTTT 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTTT 778

Qy        782 CTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGATG 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 CTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGATG 838

Qy        842 CTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTTA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 CTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTTA 898

Qy        902 GTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGTG 958

Qy        962 TAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATTT 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 TAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATTT 1018

Qy       1022 AATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTTG 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTTG 1078

Qy       1082 TTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC- 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1079 TTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATCA 1138

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1198
```

```
Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1258

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1318

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1378

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1438

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1498

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAAGATTCTCATCACTCAA 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1558

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1618

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       1619 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAAC 1678

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1738

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1798

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1858

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1918

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1978

Qy       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2038

Qy       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       2039 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTT-GTAATACCAGAACTCT 2097

Qy       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160
             ||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       2098 AATTGGA-CACCCTGAGGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2156

Qy       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2157 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2216

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2217 CGGCCAGGCACGGCGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2276

Qy       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2277 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2336

Qy       2341 ACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2337 ACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2396

Qy       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       2397 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAG---GTTGCGGTGAGCTGAGATTGCAC 2453

Qy       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCA 2503
             |||||||||||||||||||||||||||||||||||||||||||
Db       2454 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCA 2496
```

RESULT 6
AF189270

```
LOCUS       AF189270                2103 bp    mRNA    linear   PRI 22-MAR-2001
DEFINITION  Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION   AF189270
VERSION     AF189270.1  GI:10441603
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2103)
  AUTHORS   An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
            Wright,G.L. Jr. and Veltri,R.W.
  TITLE     Cloning and characterization of UROC28, a novel gene overexpressed
            in prostate, breast, and bladder cancers
  JOURNAL   Cancer Res. 60 (24), 7014-7020 (2000)
  MEDLINE   21028101
   PUBMED   11156405
REFERENCE   2  (bases 1 to 2103)
  AUTHORS   An,G. and Veltri,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
            Parkway, Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1. .2103
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q23-q24"
                     /tissue_type="prostate cancer"
     gene            1. .2103
                     /gene="UC28"
     CDS             99. .506
                     /gene="UC28"
                     /codon_start=1
                     /product="UC28 protein"
                     /protein_id="AAG17118.1"
                     /db_xref="GI:10441604"
                     /translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
                     LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     misc_feature    1961. .2103
                     /gene="UC28"
                     /note="alternative splicing"
     polyA_signal    2047. .2052
                     /gene="UC28"
ORIGIN
```

```
  Query Match             78.3%;  Score 1961.6;  DB 9;  Length 2103;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 1968;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978
               ||||||||||||||||||||||||||||||||||||||||   |  |||  ||  | |
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
```

```
RESULT 7
AR146835
LOCUS       AR146835               2087 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 83 from patent US 6218529.
ACCESSION   AR146835
VERSION     AR146835.1  GI:15110024
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2087)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 83 17-APR-2001;
```

```
FEATURES             Location/Qualifiers
     source          1. .2087
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           77.8%;  Score 1950;  DB 6;  Length 2087;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1967;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978
              |||||||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1977

RESULT 8
AR146757
LOCUS       AR146757                757 bp    DNA     linear   PAT 08-AUG-2001
```

```
DEFINITION  Sequence 3 from patent US 6218529.
ACCESSION   AR146757
VERSION     AR146757.1  GI:15109946
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 3 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             25.0%;  Score 626.2;  DB 6;  Length 757;
  Best Local Similarity   94.6%;  Pred. No. 1.6e-123;
  Matches  716;  Conservative    0;  Mismatches   28;  Indels   13;  Gaps    6;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db             1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036
             |||||||||||    | |||||||||||||||| |||||||||||| | |      |
Db           301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
             |||   | |    | |   | || |||||| |||||||||||   || ||||    |||
Db           361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db           421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660

Qy       2333 TACTAAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           661 TACTAAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429
             |||||||||||||||||||||||||||||||||||||
Db           721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757
```

```
RESULT 9
AC092933/c
LOCUS       AC092933               137227 bp    DNA     linear   PRI 26-SEP-2002
DEFINITION  Homo sapiens 3 BAC RP11-200I19 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC092933
VERSION     AC092933.21  GI:22657492
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137227)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 137227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 137227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 137227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 137227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 3, 2002 this sequence version replaced gi:22549647.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.
FEATURES             Location/Qualifiers
     source          1. .137227
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-200I19"
     misc_feature    1. .2012
                     /note="overlaps bases 45945. .47956 of clone AC117490"
                     /function="clone overlap"
     STS             17796. .17916
                     /standard_name="RH41804"
     misc_feature    19964
                     /note="Polymorphism between Baylor BACS (C/T)."
                     /function="polymorphic site"
     STS             40291. .40674
                     /standard_name="GDB:631802"
     misc_feature    42127
                     /note="Polymorphism between Baylor BACS (C/T)."
                     /function="polymorphic site"
     misc_feature    42419. .42420
                     /note="Polymorphism between Baylor BACS (AG/--)."
                     /function="polymorphic site"
     STS             46723. .46840
                     /standard_name="D10S16"
     STS             63403. .63697
                     /standard_name="SGC38304"
     STS             63466. .63604
                     /standard_name="SHGC-77695"
     STS             65307. .65468
                     /standard_name="RH99296"
     STS             94390. .94510
                     /standard_name="D3S3440"
     STS             94393. .94560
                     /standard_name="SHGC-3953"
     STS             95496. .95654
                     /standard_name="RH66305"
     STS             95498. .95656
                     /standard_name="SHGC-77694"
     STS             97480. .97732
                     /standard_name="WI-22352"
     STS             97480. .97611
                     /standard_name="SHGC-77696"
     STS             97878. .98036
                     /standard_name="NIB2025"
     STS             97910. .98094
                     /standard_name="SHGC-77693"
     STS             99577. .99720
                     /standard_name="RH102770"
     STS             104474. .104588
                     /standard_name="D10S16"
     misc_feature    135224. .137227
                     /note="overlaps bases 1. .2004 of clone AC023797"
                     /function="clone overlap"
ORIGIN

  Query Match             8.0%;  Score 201.6;  DB 9;  Length 137227;
  Best Local Similarity   85.3%;  Pred. No. 5.1e-33;
  Matches  249;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    2;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 2266
              | |||  ||       |||  |||||||||||||||| ||||||||||  ||||||||| ||
Db       4672 AAATAGCAAATGGAAGGCTGGGCACGGTGGCTCACACCTGTAATCCTGGCACTTTGGAAG 4613

Qy       2267 GTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCC 2326
              |  |||||||||||||||||| ||||||| |||||||||| ||| ||||||  ||||||
Db       4612 GCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAAACCCC 4553

Qy       2327 CATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAG 2386
              | ||||||||||||  |||||||||||| |||||||||| ||||||||||||||||||
Db       4552 CGTCTCTACTAAAATTACAAAAAATTAGCTGGGCGTGGTGGGACATGCCTGTAATCCCAG 4493

Qy       2387 CTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGT 2445
              |||||||||||||||||| |||||||||||||||||| ||||||||   | ||| ||
Db       4492 CTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGAG---GCTGCAGT 4436

Qy       2446 GAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
              | |||||||||| | |||| |||||||||||||||||| |||  |||||||
Db       4435 TACCTGAGATTGCGCCATTGTACTCCAGCCTGGGCAACAAGAGCGAAACTCC 4384


RESULT 10
AC023415
LOCUS       AC023415              166895 bp    DNA     linear   HTG 08-APR-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-200I19 map 3, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC023415
VERSION     AC023415.3  GI:7528128
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166895)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-200I19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166895)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
```

```
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 8, 2000 this sequence version replaced gi:7342076.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6745
                Center clone name: 200_I_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 154794 bases at least Q40
                Consensus quality: 160262 bases at least Q30
                Consensus quality: 162545 bases at least Q20
                Insert size: 171000; agarose-fp
                Insert size: 164295; sum-of-contigs
                Quality coverage: 4.7 in Q20 bases; agarose-fp
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1021: contig of 1021 bp in length
            *     1022     1121: gap of 100 bp
            *     1122     2188: contig of 1067 bp in length
            *     2189     2288: gap of 100 bp
            *     2289     3869: contig of 1581 bp in length
            *     3870     3969: gap of 100 bp
            *     3970     5020: contig of 1051 bp in length
            *     5021     5120: gap of 100 bp
            *     5121     6753: contig of 1633 bp in length
            *     6754     6853: gap of 100 bp
            *     6854     8166: contig of 1313 bp in length
            *     8167     8266: gap of 100 bp
            *     8267     9661: contig of 1395 bp in length
            *     9662     9761: gap of 100 bp
            *     9762    12386: contig of 2625 bp in length
            *    12387    12486: gap of 100 bp
            *    12487    15618: contig of 3132 bp in length
            *    15619    15718: gap of 100 bp
            *    15719    18825: contig of 3107 bp in length
            *    18826    18925: gap of 100 bp
            *    18926    22110: contig of 3185 bp in length
            *    22111    22210: gap of 100 bp
            *    22211    25885: contig of 3675 bp in length
            *    25886    25985: gap of 100 bp
            *    25986    31121: contig of 5136 bp in length
            *    31122    31221: gap of 100 bp
            *    31222    35458: contig of 4237 bp in length
            *    35459    35558: gap of 100 bp
            *    35559    40046: contig of 4488 bp in length
            *    40047    40146: gap of 100 bp
            *    40147    47343: contig of 7197 bp in length
            *    47344    47443: gap of 100 bp
            *    47444    52032: contig of 4589 bp in length
            *    52033    52132: gap of 100 bp
            *    52133    57816: contig of 5684 bp in length
            *    57817    57916: gap of 100 bp
            *    57917    63002: contig of 5086 bp in length
            *    63003    63102: gap of 100 bp
            *    63103    70861: contig of 7759 bp in length
            *    70862    70961: gap of 100 bp
            *    70962    76669: contig of 5708 bp in length
            *    76670    76769: gap of 100 bp
            *    76770    84103: contig of 7334 bp in length
            *    84104    84203: gap of 100 bp
            *    84204    91221: contig of 7018 bp in length
            *    91222    91321: gap of 100 bp
            *    91322   100954: contig of 9633 bp in length
            *   100955   101054: gap of 100 bp
            *   101055   109879: contig of 8825 bp in length
            *   109880   109979: gap of 100 bp
            *   109980   126962: contig of 16983 bp in length
            *   126963   127062: gap of 100 bp
            *   127063   166895: contig of 39833 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166895
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-200I19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1021
                     /note="assembly_fragment"
     misc_feature    1122. .2188
                     /note="assembly_fragment"
     misc_feature    2289. .3869
                     /note="assembly_fragment"
     misc_feature    3970. .5020
                     /note="assembly_fragment"
     misc_feature    5121. .6753
                     /note="assembly_fragment"
     misc_feature    6854. .8166
                     /note="assembly_fragment"
     misc_feature    8267. .9661
                     /note="assembly_fragment"
     misc_feature    9762. .12386
                     /note="assembly_fragment"
     misc_feature    12487. .15618
                     /note="assembly_fragment"
     misc_feature    15719. .18825
                     /note="assembly_fragment"
     misc_feature    18926. .22110
                     /note="assembly_fragment"
     misc_feature    22211. .25885
                     /note="assembly_fragment"
     misc_feature    25986. .31121
                     /note="assembly_fragment"
     misc_feature    31222. .35458
                     /note="assembly_fragment"
     misc_feature    35559. .40046
```

```
                     /note="assembly_fragment"
     misc_feature    40147. .47343
                     /note="assembly_fragment"
     misc_feature    47444. .52032
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    52133. .57816
                     /note="assembly_fragment"
     misc_feature    57917. .63002
                     /note="assembly_fragment"
     misc_feature    63103. .70861
                     /note="assembly_fragment"
     misc_feature    70962. .76669
                     /note="assembly_fragment"
     misc_feature    76770. .84103
                     /note="assembly_fragment"
     misc_feature    84204. .91221
                     /note="assembly_fragment"
     misc_feature    91322. .100954
                     /note="assembly_fragment"
     misc_feature    101055. .109879
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    109980. .126962
                     /note="assembly_fragment"
     misc_feature    127063. .166895
                     /note="assembly_fragment"
ORIGIN

  Query Match             8.0%;  Score 201.6;  DB 2;  Length 166895;
  Best Local Similarity   85.3%;  Pred. No. 5e-33;
  Matches  249;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    2;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 2266
              | |||  ||      |||  ||||||||||||||| ||||||||||  |||||||| ||
Db     105201 AAATAGCAAATGGAAGGCTGGGCACGGTGGCTCACACCTGTAATCCTGGCACTTTGGAAG 105260

Qy       2267 GTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCC 2326
              |  |||||||||||||||||| ||||||| |||||||||| ||| ||||||  ||||||
Db     105261 GCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAAACCCC 105320

Qy       2327 CATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAG 2386
              | ||||||||||||  |||||||||||||| |||||||||| ||||||||||||||||||
Db     105321 CGTCTCTACTAAAATTACAAAAAATTAGCTGGGCGTGGTGGGACATGCCTGTAATCCCAG 105380

Qy       2387 CTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGT 2445
              |||||||||||||||||| |||||||||||||||||||| ||||||||   | ||| ||
Db     105381 CTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGAG---GCTGCAGT 105437

Qy       2446 GAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
              | |||||||||| | |||| |||||||||||||||| |||  |||||||
Db     105438 TACCTGAGATTGCGCCATTGTACTCCAGCCTGGGCAACAAGAGCGAAACTCC 105489


RESULT 11
AC112202
LOCUS       AC112202               170961 bp    DNA     linear   HTG 20-FEB-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-6E10, WORKING DRAFT SEQUENCE,
            9 unordered pieces.
ACCESSION   AC112202
VERSION     AC112202.1  GI:18767474
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170961)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170961)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 408775
            Center clone name: RPCI-11_6E10
            -------
            Summary Statistics
            Consensus quality: 163237 bases at least Q40
            Consensus quality: 165500 bases at least Q30
            Consensus quality: 166457 bases at least Q20
            Estimated insert size: 179000; agarose-fp estimation
            Estimated insert size: 170161; sum-of-contigs estimation
            Quality coverage:  14.19 in Q20 bases; agarose-fp estimation
            Quality coverage:  14.93 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1074: contig of 1074 bp in length
            *     1075     1174: gap of unknown length
            *     1175     2298: contig of 1124 bp in length
            *     2299     2398: gap of unknown length
            *     2399     3709: contig of 1311 bp in length
            *     3710     3809: gap of unknown length
            *     3810     5697: contig of 1888 bp in length
            *     5698     5797: gap of unknown length
            *     5798     7062: contig of 1265 bp in length
            *     7063     7162: gap of unknown length
            *     7163    10130: contig of 2968 bp in length
            *    10131    10230: gap of unknown length
            *    10231    37572: contig of 27342 bp in length
            *    37573    37672: gap of unknown length
            *    37673    78655: contig of 40983 bp in length
            *    78656    78755: gap of unknown length
            *    78756   170961: contig of 92206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170961
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-6E10"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             8.0%;  Score 201.6;  DB 2;  Length 170961;
  Best Local Similarity   85.3%;  Pred. No. 5e-33;
  Matches  249;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    2;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 2266
              | |||  ||      |||  ||||||||||||||| ||||||||||  |||||||| ||
Db     93321 AAATAGCAAATGGAAGGCTGGGCACGGTGGCTCACACCTGTAATCCTGGCACTTTGGAAG 93380

Qy       2267 GTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCC 2326
              |  |||||||||||||||||| ||||||| |||||||||| ||| ||||||  ||||||
Db     93381 GCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAAACCCC 93440

Qy       2327 CATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAG 2386
              | ||||||||||||  |||||||||||||| |||||||||| ||||||||||||||||||
Db     93441 CGTCTCTACTAAAATTACAAAAAATTAGCTGGGCGTGGTGGGACATGCCTGTAATCCCAG 93500

Qy       2387 CTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGT 2445
```

```
                ||||||||||||||||||| |||||||||||||||||||| ||||||||   | ||| ||
Db      93501 CTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGAG---GCTGCAGT 93557

Qy       2446 GAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
                | |||||||||| | |||| |||||||||||||||||| |||  |||||||
Db      93558 TACCTGAGATTGCGCCATTGTACTCCAGCCTGGGCAACAAGAGCGAAACTCC 93609


RESULT 12
AC131148/c
LOCUS       AC131148              174054 bp    DNA     linear   HTG 21-SEP-2002
DEFINITION  Homo sapiens clone RP11-622J3, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   AC131148
VERSION     AC131148.2  GI:23264913
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174054)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174054)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 174054)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 21, 2002 this sequence version replaced gi:22296651.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HEKP
                Center clone name: RP11-622J3
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 170243 bases at least Q40
                Consensus quality: 170621 bases at least Q30
                Consensus quality: 170908 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4685: contig of 4685 bp in length
            *     4686     4785: gap of unknown length
            *     4786     9848: contig of 5063 bp in length
            *     9849     9948: gap of unknown length
            *     9949    30170: contig of 20222 bp in length
            *    30171    30270: gap of unknown length
            *    30271    51230: contig of 20960 bp in length
            *    51231    51330: gap of unknown length
            *    51331    76881: contig of 25551 bp in length
            *    76882    76981: gap of unknown length
            *    76982   174054: contig of 97073 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174054
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-622J3"
ORIGIN

  Query Match             8.0%;  Score 201.6;  DB 2;  Length 174054;
  Best Local Similarity   85.3%;  Pred. No. 5e-33;
  Matches  249;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    2;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 2266
                | |||  ||      |||  |||||||||||||| ||||||||| |||||||| ||
Db      14631 AAATAGCAAATGGAAGGCTGGGCACGGTGGCTCACACCTGTAATCCTGGCACTTTGGAAG 14572

Qy       2267 GTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCC 2326
                |  |||||||||||||||||| ||||||| |||||||||| ||| ||||||  ||||||
Db      14571 GCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAAACCCC 14512

Qy       2327 CATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAG 2386
                | ||||||||||||  |||||||||||| |||||||||| ||||||||||||||||||
Db      14511 CGTCTCTACTAAAATTACAAAAAATTAGCTGGGCGTGGTGGGACATGCCTGTAATCCCAG 14452

Qy       2387 CTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGT 2445
                |||||||||||||||||| |||||||||||||||||| ||||||||   | ||| ||
Db      14451 CTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGAG---GCTGCAGT 14395

Qy       2446 GAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
                | |||||||||| | |||| |||||||||||||||||| |||  |||||||
Db      14394 TACCTGAGATTGCGCCATTGTACTCCAGCCTGGGCAACAAGAGCGAAACTCC 14343
```

```
RESULT 13
AF168787
LOCUS       AF168787               200237 bp    DNA     linear   PRI 14-MAR-2000
DEFINITION  Homo sapiens vanilloid receptor gene, partial sequence; CARKL and
            CTNS genes, complete cds; TIP1 gene, partial cds; P2X5b and P2X5a
            genes, complete cds; and HUMINAE gene, partial cds.
ACCESSION   AF168787
VERSION     AF168787.1  GI:7239175
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200237)
  AUTHORS   Touchman,J.W., Anikster,Y., Dietrich,N.L., Maduro,V.V.,
            McDowell,G., Shotelersuk,V., Bouffard,G.G.,
            Beckstrom-Sternberg,S.M., Gahl,W.A. and Green,E.D.
  TITLE     The genomic region encompassing the nephropathic cystinosis gene
            (CTNS): complete sequencing of a 200-kb segment and discovery of a
            novel gene within the common cystinosis-causing deletion
  JOURNAL   Genome Res. 10 (2), 165-173 (2000)
  MEDLINE   20138496
   PUBMED   10673275
REFERENCE   2  (bases 1 to 200237)
  AUTHORS   Dietrich,N.L., Fergusson,M., Sorbello,E., Gupta,A., Torkzadeh,R.,
            Varner,C., Walker,M., Bouffard,G.G., Beckstrom-Sternberg,S.M. and
            Touchman,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1999) Intramural Sequencing Center, National
            Institutes of Health, 8717 Grovemont Circle, Gaithersburg, Maryland
            20877, USA
FEATURES             Location/Qualifiers
     source          1. .200237
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p13"
     gene            complement(<1. .27738)
                     /gene="vanilloid receptor"
                     /note="coding sequence not determined; similar to Rattus
                     norvegicus vanilloid receptor subtype 1 encoded by GenBank
                     Accession Number AF029310"
                     /evidence=not_experimental
     repeat_region   5. .79
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   129. .314
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   1252. .1851
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   2437. .2736
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   3093. .3142
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   3389. .3685
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   4117. .4246
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   4254. .4625
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   4616. .4784
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   5466. .5735
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(6071. .6363)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   7935. .8018
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   8055. .8343
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   10007. .10275
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   10255. .10507
                     /rpt_family="MER42"
                     /rpt_type=dispersed
     repeat_region   10549. .10838
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   10834. .11217
                     /rpt_family="MER42"
                     /rpt_type=dispersed
     repeat_region   11289. .12215
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   13386. .13659
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   13735. .13906
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   14075. .14295
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   complement(14848. .15120)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   15173. .15439
                     /rpt_family="MER20"
                     /rpt_type=dispersed
     repeat_region   16062. .16377
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(16495. .16633)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   16960. .17705
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   17979. .18328
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   19355. .19786
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   19869. .20162
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(20503. .20804)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   21844. .22129
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   22918. .23184
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   23880. .24184
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   24368. .24421
                     /rpt_family="Alu"
```

```
                /rpt_type=dispersed
     repeat_region   complement(24564. .24833)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   complement(28469. .28625)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   29167. .29464
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   complement(30577. .30675)
                /rpt_family="MIR"
                /rpt_type=dispersed
     repeat_region   complement(32993. .33290)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   complement(33885. .34180)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   34276. .34562
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   complement(34947. .35225)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   35626. .35919
                /rpt_family="Alu"
                /rpt_type=dispersed
     misc_recomb     36253
                /note="cystinosis deletion breakpoint"
                /evidence=experimental
     repeat_region   complement(36925. .37208)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   38170. .38470
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   39628. .39934
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   40422. .40545
                /rpt_family="MIR"
                /rpt_type=dispersed
     repeat_region   40570. .40839
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   complement(41304. .41558)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   42368. .42675
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   complement(42857. .43141)
                /rpt_family="Alu"
                /rpt_type=dispersed
     repeat_region   43262. .43546
                /rpt_family="Alu"

  Query Match             7.9%;  Score 198.8;  DB 9;  Length 200237;
  Best Local Similarity   87.5%;  Pred. No. 1.9e-32;
  Matches  266;  Conservative    0;  Mismatches   27;  Indels   11;  Gaps    4;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGG--- 2263
             | || |||| ||||| ||| |||| ||||||||||||||||||||||||||||||||
Db     137661 ACATTAAAAAACCCCAGCCGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 137720

Qy       2264 ---GAGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGG 2320
                |||||||||||||||||||||||| |||||||||||||||||||||| ||||| |
Db     137721 GCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGG-AG 137779

Qy       2321 AACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAA 2380
              || |||||||||||||||||| |||||||| ||||||| ||||||||||||||||||||
Db     137780 AAACCCCATCTCTACTAAAAATACAAAAAAGTAGCCGGACGTGGTGGCACATGCCTGTAA 137839

Qy       2381 TCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGT 2439
              |||||||||||| |||||||||||| |||||||||||||||||| | |||  ||||||||
Db     137840 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCAGAGGTTGT 137899

Qy       2440 TGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCC 2499
                |||||||||||||  | | |||||||||||||||||||||| ||| |||||||  |
Db     137900 ---GGTGAGCTGAGATGACGCCATTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCTGC 137956

Qy       2500 CCCA 2503
              | ||
Db     137957 CTCA 137960


RESULT 14
AC009301/c
LOCUS       AC009301               129625 bp    DNA     linear   PRI 30-SEP-2000
DEFINITION  Homo sapiens BAC clone RP11-62F14 from 2, complete sequence.
ACCESSION   AC009301
VERSION     AC009301.3  GI:7243885
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129625)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 129625)
  AUTHORS   Drone,K., Leonard,S. and Doebber,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-62F14
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 129625)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 129625)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 129625)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 15, 2000 this sequence version replaced gi:6598926.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0062F14
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
```

restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-153B21, 200 bp overlap. Actual start of this clone is at base position 133171 of RP11-153B21; actual end is at base position 129625 of RP11-62F14.

FEATURES Location/Qualifiers

source	1. .129625	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2"
		/clone="RP11-62F14"
		/clone_lib="RPCI-11"
misc_feature	1539. .1647	/note="match to EST AA358669 (NID:g2010987)"
repeat_region	1752. .1921	/rpt_family="MIR"
repeat_region	2556. .2676	/rpt_family="MIR"
repeat_region	3904. .4038	/rpt_family="MIR"
repeat_region	4064. .4161	/rpt_family="MIR"
repeat_region	5485. .5701	/rpt_family="MIR"
repeat_region	5766. .5858	/rpt_family="MER1_type"
repeat_region	5859. .6170	/rpt_family="Alu"
repeat_region	6171. .6270	/rpt_family="MER1_type"
repeat_region	6302. .6394	/rpt_family="L2"
repeat_region	6369. .6482	/rpt_family="MIR"
misc_feature	7024. .7525	/note="match to EST AA179680 (NID:g1761913) zp13e05.r1"
misc_feature	7845. .7846	/note="match to EST AA176625 (NID:g1757748) zp13e05.s1"
misc_feature	7931. .8410	/note="match to EST AA176625 (NID:g1757748) zp13e05.s1"
repeat_region	8170. .8362	/rpt_family="MIR"
repeat_region	8405. .8452	/rpt_family="L2"
repeat_region	8428. .8707	/rpt_family="L2"
repeat_region	8925. .9219	/rpt_family="L2"
misc_feature	9387. .9454	/note="match to EST AA358669 (NID:g2010987)"
repeat_region	10445. .10657	/rpt_family="MIR"
repeat_region	11048. .11344	/rpt_family="Alu"
repeat_region	11348. .11496	/rpt_family="L1"
repeat_region	11498. .11520	/rpt_family="AT_rich"
misc_feature	11874. .11941	/note="match to EST AA358669 (NID:g2010987)"
misc_feature	12150. .12224	/note="match to EST AA285241 (NID:g1929551)"
repeat_region	12358. .12434	/rpt_family="MIR"
repeat_region	12805. .12833	/rpt_family="AT_rich"
repeat_region	12939. .13014	/rpt_family="MIR"
repeat_region	13382. .13500	/rpt_family="Alu"
repeat_region	13603. .13939	/rpt_family="L2"
repeat_region	14032. .14081	/rpt_family="L2"
repeat_region	14244. .14554	/rpt_family="Alu"
repeat_region	15087. .15575	/rpt_family="L1"
repeat_region	16160. .16221	/rpt_family="(TG)n"
repeat_region	16224. .16856	/rpt_family="L1"
repeat_region	17784. .17926	/rpt_family="MIR"
repeat_region	19043. .19564	/rpt_family="MER21-group"
repeat_region	20471. .21063	/rpt_family="L2"
repeat_region	21165. .21598	/rpt_family="MaLR"
misc_feature	21972. .22068	/note="match to EST AA285241 (NID:g1929551)"
misc_feature	22909. .22993	/note="match to EST AA285241 (NID:g1929551)"
misc_feature	22935. .22993	/note="match to EST AA081997 (NID:g1624063) zn26d08.r1"
repeat_region	24647. .25035	/rpt_family="MaLR"
misc_feature	25547. .25650	/note="match to EST AA081997 (NID:g1624063) zn26d08.r1"
misc_feature	25547. .25576	/note="match to EST AA285241 (NID:g1929551)"
repeat_region	26490. .26608	/rpt_family="L2"
repeat_region	29443. .29466	/rpt_family="(GA)n"
repeat_region	29466. .29491	/rpt_family="(TG)n"
repeat_region	29522. .29561	/rpt_family="(TG)n"
repeat_region	29899. .29989	/rpt_family="MIR"
repeat_region	30000. .30490	/rpt_family="MaLR"
repeat_region	33975. .34081	/rpt_family="MIR"
repeat_region	34579. .34750	/rpt_family="MER1_type?"
STS	35327. .35502	/db_xref="GI:4529205"
repeat_region	35460. .35512	/rpt_family="MIR"
misc_feature	35657. .35783	/note="match to EST AA081997 (NID:g1624063) zn26d08.r1"

Query Match 7.9%; Score 198.6; DB 9; Length 129625;
Best Local Similarity 83.2%; Pred. No. 2.3e-32;

```
  Matches  252;  Conservative    0;  Mismatches   44;  Indels    7;  Gaps    2;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG-- 2264
              || |||      |   |||| ||  | ||| | || ||||||||||||||||||||||
Db     104230 ATTTAATGGCTCAAAGGCCGGGAGCAGTGACCCATGCCTGTAATCCCAGCACTTTGGGAG 104171

Qy       2265 ----AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGG 2320
                |||||||  ||||||||||||||| |||||||||||||||||||| | | |||||||
Db     104170 GCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCTAACATGGTGA 104111

Qy       2321 AACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAA 2380
              ||| ||||| |||||||||| |||||||||||||| |||||||||||||| ||||||||
Db     104110 AAC-CCCATTTCTACTAAAAATACAAAAAATTAGCCAGGCGTGGTGGCACACGCCTGTAA 104052

Qy       2381 TCCCAGCTACTCAGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTT 2440
              || |||||||||||||||||| ||||||| ||||||||| || ||||||||||||
Db     104051 TCTCAGCTACTCAGGAGGCTGAGGAGGAGAATCGCTTGAACCTGGGAGGTGGAGGTTGCA 103992

Qy       2441 GCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCC 2500
              || ||||||||||||||  | |||||||||||||||||||||| |||  |||||||  |
Db     103991 GCAGTGAGCTGAGATTGTGCCATTGCACTCCAGCCTGGGCAACAAGAGAGAAACTCCATC 103932

Qy       2501 CCA 2503
              ||
Db     103931 TCA 103929


RESULT 15
AC107422
LOCUS       AC107422               318586 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-372J10, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC107422
VERSION     AC107422.3  GI:20335954
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 318586)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 318586)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 318586)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2002 this sequence version replaced gi:18449820.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HDRW
                Center clone name: RP11-372J10
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 306871 bases at least Q40
                Consensus quality: 310966 bases at least Q30
                Consensus quality: 313294 bases at least Q20
                Estimated insert size: 170415; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4906: contig of 4906 bp in length
            *     4907     5006: gap of unknown length
            *     5007    12370: contig of 7364 bp in length
            *    12371    12470: gap of unknown length
            *    12471    21459: contig of 8989 bp in length
            *    21460    21559: gap of unknown length
            *    21560    33615: contig of 12056 bp in length
            *    33616    33715: gap of unknown length
            *    33716    47492: contig of 13777 bp in length
            *    47493    47592: gap of unknown length
            *    47593    69615: contig of 22023 bp in length
            *    69616    69715: gap of unknown length
            *    69716    95747: contig of 26032 bp in length
            *    95748    95847: gap of unknown length
            *    95848   121218: contig of 25371 bp in length
            *   121219   121318: gap of unknown length
            *   121319   146590: contig of 25272 bp in length
            *   146591   146690: gap of unknown length
            *   146691   168128: contig of 21438 bp in length
            *   168129   168228: gap of unknown length
            *   168229   194927: contig of 26699 bp in length
            *   194928   195027: gap of unknown length
            *   195028   221744: contig of 26717 bp in length
```

```
            *   221745   221844: gap of unknown length
            *   221845   267827: contig of 45983 bp in length
            *   267828   267927: gap of unknown length
            *   267928   318586: contig of 50659 bp in length.
FEATURES             Location/Qualifiers
     source          1. .318586
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-372J10"
ORIGIN

  Query Match             7.9%;  Score 198.6;  DB 2;  Length 318586;
  Best Local Similarity   83.2%;  Pred. No. 2e-32;
  Matches  252;  Conservative    0;  Mismatches   44;  Indels    7;  Gaps    2;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG-- 2264
              || |||     |   |||| ||  | ||| | || ||||||||||||||||||||||
Db     242665 ATTTAATGGCTCAAAGGCCGGGAGCAGTGACCCATGCCTGTAATCCCAGCACTTTGGGAG 242724

Qy       2265 ----AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGG 2320
                |||||||  ||||||||||||||| |||||||||||||||||| | | |||||||
Db     242725 GCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCTAACATGGTGA 242784

Qy       2321 AACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAA 2380
              ||| ||||| |||||||||| |||||||||||||| |||||||||||||| ||||||||
Db     242785 AAC-CCCATTTCTACTAAAAATACAAAAAATTAGCCAGGCGTGGTGGCACACGCCTGTAA 242843

Qy       2381 TCCCAGCTACTCAGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTT 2440
              || ||||||||||||||||||| ||||||| ||||||||| || ||||||||||||
Db     242844 TCTCAGCTACTCAGGAGGCTGAGGAGGAGAATCGCTTGAACCTGGGAGGTGGAGGTTGCA 242903

Qy       2441 GCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCC 2500
              || ||||||||||||||  | |||||||||||||||||||||| |||  ||||||| |
Db     242904 GCAGTGAGCTGAGATTGTGCCATTGCACTCCAGCCTGGGCAACAAGAGAGAAACTCCATC 242963

Qy       2501 CCA 2503
                ||
Db     242964 TCA 242966


Search completed: September  9, 2005, 03:30:41
Job time : 7229.71 secs
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:23:54 ; Search time 886.428 Seconds
(without alignments)
16735.564 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2505.6	100.0	2506	3	AAZ87584	Aaz87584 Prostate
	2	2494	99.5	2505	4	AAS04001	Aas04001 Biomarker
	3	1961.6	78.3	2088	3	AAZ87583	Aaz87583 Prostate
	4	1950	77.8	2087	4	AAS04000	Aas04000 Biomarker
	5	626.2	25.0	757	2	AAV16883	Aav16883 Human pro
	6	626.2	25.0	757	2	AAX26018	Aax26018 Prostate
	7	626.2	25.0	757	3	AAZ87503	Aaz87503 Prostate,
	8	626.2	25.0	757	4	AAS03722	Aas03722 Biomarker
c	9	190.6	7.6	299598	12	ADQ59380	Adq59380 Human can
c	10	186.8	7.5	215974	12	ADQ97523	Adq97523 Human can
	11	186.4	7.4	2680	4	AAK85292	Aak85292 Human imm
c	12	186.4	7.4	5788	5	ABA15481	Aba15481 Human ner
c	13	186.4	7.4	5788	5	ABA15482	Aba15482 Human ner
c	14	186.4	7.4	14417	4	AAK87584	Aak87584 Human imm
	15	186.4	7.4	14417	4	AAK71816	Aak71816 Human imm
c	16	186.4	7.4	14417	4	AAK73113	Aak73113 Human imm
	17	186.4	7.4	14417	4	AAI62923	Aai62923 Human gen
c	18	186.4	7.4	14417	8	ABZ74674	Abz74674 Secreted
c	19	186.4	7.4	14417	10	ABZ68196	Abz68196 Human sec
c	20	186.4	7.4	14426	4	AAK73099	Aak73099 Human imm
c	21	186.4	7.4	14426	4	AAK87568	Aak87568 Human imm
	22	186.4	7.4	14426	4	AAK85290	Aak85290 Human imm
	23	186.4	7.4	14426	4	AAK71814	Aak71814 Human imm
	24	186.4	7.4	14426	4	AAI62921	Aai62921 Human gen
c	25	186.4	7.4	14426	8	ABZ74673	Abz74673 Secreted
c	26	186.4	7.4	14426	10	ABZ68195	Abz68195 Human sec
	27	184.4	7.4	874	4	AAH98452	Aah98452 Human EST
	28	184.4	7.4	1599	10	ADG10849	Adg10849 Human STA
c	29	183.8	7.3	41150	10	ADL13819	Adl13819 Osteoarth
	30	183.8	7.3	44348	12	ADN48556	Adn48556 Human Not
	31	183.6	7.3	106746	3	AAA10225	Aaa10225 Human PCT
c	32	183.6	7.3	108316	10	ADC87336	Adc87336 Human GPC
c	33	183.4	7.3	5286	13	ADS89008	Ads89008 Human GGT
c	34	183.2	7.3	8846	4	ABA07221	Aba07221 Human pan
c	35	183.2	7.3	8846	4	AAK89858	Aak89858 Human dig
	36	183.2	7.3	110665	12	ADQ19183	Adq19183 Human sof
	37	183.2	7.3	227968	6	ABK83497	Abk83497 Human cDN
	38	183.2	7.3	227968	12	ADQ18538	Adq18538 Human sof
	39	183	7.3	2952	6	ABK69090	Abk69090 DNA encod
c	40	183	7.3	73723	13	ABD33145	Abd33145 Human can
	41	183	7.3	270150	11	ADP65796	Adp65796 Human 16p
	42	182.8	7.3	463	9	ACH36634	Ach36634 Human end
	43	182.6	7.3	89625	11	ACN45194	Acn45194 Human gen
c	44	182.2	7.3	71048	11	ACN44082	Acn44082 Human gen
c	45	181.8	7.3	91760	11	ACN44410	Acn44410 Human gen

ALIGNMENTS

RESULT 1
AAZ87584
ID AAZ87584 standard; DNA; 2506 BP.
XX
AC AAZ87584;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 splice variant.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59296.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Claim 1; Page 184-186; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid

CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Query Match 100.0%; Score 2505.6; DB 3; Length 2506;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980

Qy       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040

Qy       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100

Qy       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160

Qy       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280

Qy       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340

Qy       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400

Qy       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460

Qy       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
                ||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506


RESULT 2
AAS04001
ID   AAS04001 standard; cDNA; 2505 BP.
XX
AC   AAS04001;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             99. .506
FT                   /*tag=  a
FT                   /product= "Prostate cancer marker protein"
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   P-PSDB; AAU02175.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Claim 2; Col 121-125; 78pp; English.
XX
CC   The sequence represents nucleic acid biomarker UC band 28 #3, used in
CC   detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC   sequences can be used as hybridisation probes and primers that
CC   specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC   (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC   nucleic acid markers can be used to produce antibodies for the detection
CC   of prostate, breast or bladder cancer. The nucleic acids can be used as
CC   targets for therapeutic intervention in these diseases, in the
CC   identification and isolation of full-length gene sequences, including
CC   regulatory elements for gene expression, from genomic human DNA
CC   libraries, as hybridisation probes for screening genomic human DNA
CC   libraries. The kits comprising the nucleic acid sequences are useful for
CC   detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ   Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;

  Query Match             99.5%;  Score 2494;  DB 4;  Length 2505;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2505;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
```

```
Qy       661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
Db       661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy       721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
Db       721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy       781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
Db       781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
Db      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
Db      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
Db      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
Db      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
Db      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
Db      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
Db      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
Db      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
Db      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy      1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
Db      1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy      1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
Db      1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy      1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
Db      1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy      1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
Db      1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy      1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980
Db      1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1979

Qy      1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040
Db      1980 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2039

Qy      2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100
Db      2040 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2099

Qy      2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160
Db      2100 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2159

Qy      2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220
Db      2160 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2219

Qy      2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280
Db      2220 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2279

Qy      2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
Db      2280 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2339

Qy      2341 ACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
Db      2340 ACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2399

Qy      2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
Db      2400 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2459

Qy      2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
Db      2460 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2505
```

RESULT 3
AAZ87583
ID AAZ87583 standard; DNA; 2088 BP.
XX
AC AAZ87583;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX

PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59295.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Claim 1; Page 182-183; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
SQ Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;

Query Match 78.3%; Score 1961.6; DB 3; Length 2088;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978
             ||||||||||||||||||||||||||||||||||||||    |  |||  ||  | |
Db       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1978
```

```
RESULT 4
AAS04000
ID   AAS04000 standard; cDNA; 2087 BP.
XX
AC   AAS04000;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             99. .506
FT                   /*tag=  a
FT                   /product= "Prostate cancer marker protein"
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   P-PSDB; AAU02174.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Claim 2; Col 117-121; 78pp; English.
XX
CC   The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC   detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC   sequences can be used as hybridisation probes and primers that
CC   specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC   (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC   nucleic acid markers can be used to produce antibodies for the detection
CC   of prostate, breast or bladder cancer. The nucleic acids can be used as
CC   targets for therapeutic intervention in these diseases, in the
CC   identification and isolation of full-length gene sequences, including
CC   regulatory elements for gene expression, from genomic human DNA
CC   libraries, as hybridisation probes for screening genomic human DNA
CC   libraries. The kits comprising the nucleic acid sequences are useful for
CC   detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ   Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;

  Query Match             77.8%;  Score 1950;  DB 4;  Length 2087;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1967;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
```

```
Qy       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
Db      1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
Db      1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
Db      1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
Db      1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
Db      1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
Db      1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
Db      1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
Db      1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
Db      1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy      1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
Db      1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy      1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
Db      1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy      1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
Db      1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy      1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
Db      1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy      1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978
Db      1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1977
```

RESULT 5
AAV16883
ID AAV16883 standard; DNA; 757 BP.
XX
AC AAV16883;
XX
DT 07-AUG-1998 (first entry)
XX
DE Human prostate cancer marker UC Band #28.
XX
KW Prostate cancer; human; marker; diagnosis; treatment; probe; ss.
XX
OS Homo sapiens.
XX
PN WO9804689-A1.
XX
PD 05-FEB-1998.
XX
PF 31-JUL-1996; 96WO-US012516.
XX
PR 31-JUL-1996; 96WO-US012516.
XX
PA (UROC-) UROCOR INC.
XX
PI Veltri R, Ohara SM, An G, Ralph D;
XX
DR WPI; 1998-130681/12.
XX
PT Human prostate cancer marker - useful for detection and treatment of
PT human prostate cancer.
XX
PS Claim 1; Page 133-134; 229pp; English.
XX
CC This represents a marker sequence for human prostate cancer. Isolated
CC nucleic acid segments shown in AAV16881 to AAV16885, AAV16890 to
CC AAV16903, AAV26351 and AAV26352 which can act as human prostate cancer
CC markers are provided in the specification. It also provides methods for
CC identifying markers for human prostate cancer and for detection of
CC prostate cancer cells. The markers can be identified by amplifying human
CC prostate RNA to provide nucleic acid amplification products, separating
CC the products and identifying those RNA that are differentially expressed
CC between human prostate cancers versus normal or benign human prostate.
CC Prostate cancer cells in a sample can be detected by detecting a nucleic
CC acid in a sample, the nucleic acid being a prostate cancer marker.
CC Primers and probes derived from this marker can be used for the detection
CC of prostate cancer cells in a sample. Antibodies against the protein
CC encoded by the marker nucleic acid fragments, inhibitors of the protein
CC and oligonucleotides antisense to the markers can be used in the
CC treatment of prostate cancer. The antibodies can also be used for the
CC diagnosis of human prostate cancer
XX
SQ Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

Query Match 25.0%; Score 626.2; DB 2; Length 757;
Best Local Similarity 94.6%; Pred. No. 1.3e-135;
Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;

```
Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180
```

```
Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036
             |||||||||||    | |||||||||||||| |||||||||||| | |         |
Db        301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
             |||   | |     | |   | || ||||||| ||||||||||   || ||||   |||
Db        361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660

Qy       2333 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429
             |||||||||||||||||||||||||||||||||||||
Db        721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757
```

RESULT 6
AAX26018
ID AAX26018 standard; DNA; 757 BP.
XX
AC AAX26018;
XX
DT 20-MAY-1999 (first entry)
XX
DE Prostate disease marker gene fragment UC Band #28.
XX
KW Prostate cancer; benign prostatic hyperplasia; marker gene; tumour;
KW differentiation; Reverse Transcription Polymerase Chain Reaction;
KW diagnostic; progression; cancer; metastasis; human; RT-PCR; ss.
XX
OS Homo sapiens.
XX
PN US5882864-A.
XX
PD 16-MAR-1999.
XX
PF 31-JUL-1996; 96US-00692787.
XX
PR 31-JUL-1995; 95US-0001655P.
XX
PA (UROC-) UROCOR INC.
XX
PI Veltri R, Ralph D, An G, O'hara SM;
XX
DR WPI; 1999-214055/18.
XX
PT Diagnosing prostate cancer and benign prostatic hyperplasia cells - using
PT oligonucleotide probes specific for marker genes associated with tumor
PT differentiation and progression in Reverse Transcription Polymerase Chain
PT Reaction analysis.
XX
PS Claim 1; Col 71-72; 74pp; English.
XX
CC The invention relates to methods for diagnosing prostate cancer or benign
CC prostatic hyperplasia cells in a biological sample. The method uses
CC oligonucleotide probes specific for marker genes associated with tumour
CC differentiation and progression in Reverse Transcription Polymerase Chain
CC Reaction (RT-PCR) analysis. The methods are diagnostic techniques useful
CC for detecting and monitoring the progression of benign prostatic
CC hyperplasia and human prostate cancer (the most prevalent form of cancer
CC and a major cause of death in males) prior to the tumor undergoing
CC metastasis, therefore allowing the optimal method of treatment to be
CC determined before the condition becomes life threatening. The present
CC sequence represents a claimed marker gene fragment
XX
SQ Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

Query Match 25.0%; Score 626.2; DB 2; Length 757;
Best Local Similarity 94.6%; Pred. No. 1.3e-135;
Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;

```
Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036
             |||||||||||    | |||||||||||||| |||||||||||| | |         |
Db        301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
             |||   | |     | |   | || ||||||| ||||||||||   || ||||   |||
Db        361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660

Qy       2333 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429
             |||||||||||||||||||||||||||||||||||||
```

```
Db        721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757


RESULT 7
AAZ87503
ID   AAZ87503 standard; cDNA; 757 BP.
XX
AC   AAZ87503;
XX
DT   19-APR-2000  (first entry)
XX
DE   Prostate, breast and bladder cancers detecting biomarker UC Band #28.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW   diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Claim 1; Page 165; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

  Query Match             25.0%;  Score 626.2;  DB 3;  Length 757;
  Best Local Similarity   94.6%;  Pred. No. 1.3e-135;
  Matches  716;  Conservative    0;  Mismatches   28;  Indels   13;  Gaps    6;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036
              |||||||||||    | |||||||||||||| |||||||||||| | |      |
Db        301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
              |||   | |   | |   | || ||||||| ||||||||||||   || ||||    |||
Db        361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660

Qy       2333 TACTAAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TACTAAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429
              |||||||||||||||||||||||||||||||||||||
Db        721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757


RESULT 8
AAS03722
ID   AAS03722 standard; cDNA; 757 BP.
XX
AC   AAS03722;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28 #1, used in diagnosis and prognosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
XX
```

```
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Claim 2; Col 79; 78pp; English.
XX
CC   The sequence represents nucleic acid biomarker UC band 28 #1, used in
CC   detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC   sequences can be used as hybridisation probes and primers that
CC   specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC   (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC   nucleic acid markers can be used to produce antibodies for the detection
CC   of prostate, breast or bladder cancer. The nucleic acids can be used as
CC   targets for therapeutic intervention in these diseases, in the
CC   identification and isolation of full-length gene sequences, including
CC   regulatory elements for gene expression, from genomic human DNA
CC   libraries, as hybridisation probes for screening genomic human DNA
CC   libraries. The kits comprising the nucleic acid sequences are useful for
CC   detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ   Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

  Query Match             25.0%;  Score 626.2;  DB 4;  Length 757;
  Best Local Similarity   94.6%;  Pred. No. 1.3e-135;
  Matches  716;  Conservative    0;  Mismatches   28;  Indels   13;  Gaps    6;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db             1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036
             |||||||||||    | |||||||||||||| |||||||||||| | |       |
Db           301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
             |||   | |    | |   | || ||||||| ||||||||||   || ||||    |||
Db           361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152
             ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db           421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660

Qy       2333 TACTAAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           661 TACTAAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429
             |||||||||||||||||||||||||||||||||||||
Db           721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757


RESULT 9
ADQ59380/c
ID   ADQ59380 standard; DNA; 299598 BP.
XX
AC   ADQ59380;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene sequence SEQ ID NO:16.
XX
KW   human; cancer-associated gene; cancer-associated protein; cytostatic;
KW   gene therapy; vaccine; tyrosine kinase antagonist;
KW   G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004058288-A1.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040082.
XX
PR   17-DEC-2002; 2002US-00322696.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543349/52.
DR   P-PSDB; ADQ59382.
XX
PT   New cancer-associated nucleic acid for diagnosing, preventing or treating
PT   cancer (e.g. lymphoma) or for screening agents that may be used for
PT   treating or preventing cancer.
XX
PS   Claim 16; SEQ ID NO 16; 143pp; English.
XX
CC   The present invention describes human cancer-associated (CA) nucleotide
CC   sequences (I). Also described: (1) an expression vector comprising (I);
CC   (2) a host cell comprising (I) or the expression vector; (3) a microarray
CC   for detecting a CA nucleic acid; (4) an isolated polypeptide encoded
CC   within an open reading frame of a CA sequence; (5) an isolated antibody,
CC   or its antigen binding fragment, that binds to the above polypeptide; (6)
CC   a hybridoma that produces the monoclonal antibody described above; (7) a
CC   pharmaceutical composition comprising the antibody and a pharmaceutical
CC   excipient; (8) a kit for detecting or diagnosing cancer cells, comprising
CC   the above (monoclonal) antibody or polynucleotide that selectively
CC   hybridises to any of the polynucleotide sequences mentioned above; (9)
CC   methods for diagnosing cancer or for detecting the presence or absence of
CC   cancer cells in an individual; (10) a method for inhibiting growth of
CC   cancer cells in an individual; (11) a method for delivering a therapeutic
CC   agent to cancer cells in an individual; (12) an electronic library
CC   comprising the polynucleotide or polypeptide, or their fragments,
CC   mentioned above; (13) a method of screening for anticancer activity; (14)
CC   methods for detecting cancer associated with expression of a polypeptide
CC   or the presence of the antibody in a test cell or serum sample; (15) a
CC   method for screening for a bioactive agent capable of modulating the
CC   activity of a CA protein encoded by the above nucleic acid molecule; and
CC   (16) a method for treating cancers. (I) has cytostatic activity, and can
CC   be used in gene therapy, in vaccines, as a tyrosine kinase antagonist,
CC   and as a G-protein coupled receptor antagonist. The compositions and
CC   methods of the present invention can be used for diagnosing, preventing
CC   and treating cancer, especially lymphomas. They may also be used in
CC   screening for agents that may be used for treating or preventing cancer.
CC   The present sequence represents a human CA gene sequence, which is given
CC   in the exemplification of the present invention. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
```

```
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 299598 BP; 92218 A; 54563 C; 55801 G; 95585 T; 0 U; 1431 Other;

  Query Match             7.6%;  Score 190.6;  DB 12;  Length 299598;
  Best Local Similarity   78.4%;  Pred. No. 2.1e-33;
  Matches  269;  Conservative    0;  Mismatches   64;  Indels   10;  Gaps    3;

Qy       2168 ATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCCCGGCCAG 2227
             || | | |  |||    | || || |  | |      | | || | | | |   ||||||
Db     241337 ATGTATTTTATTTTATCATATATTCATAATGGAGACTTAAAATTATATTGCTTAGGCCAG 241278

Qy       2228 GCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGT------GGGTGGATCACCT 2281
             |   ||||||||||||||||||||||||||||||||||||||       ||||||||||||
Db     241277 GTGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGAGGGTGGATCACCT 241218

Qy       2282 GAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAAA 2341
             |||||| ||||||| |||||||||||||| ||||||| |||  ||||||||||||||||
Db     241217 GAGGTCAGGAGTTCTAGACCAGCCTGACCAACATGGTGAAACTTCCATCTCTACTAAAAA 241158

Qy       2342 CACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCTG 2401
             |||||||||||||||| |||||||    |||||||||||||||||||||||||||||||
Db     241157 TACAAAAAAATTAGCCGGGTGTGGTGGTGGGTGCCTGTAATCCCAGCTACTCAGGAGGCTG 241098

Qy       2402 AGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
             |||| |||||||  ||||||||  | ||   |||   ||| | ||||||||||| || |
Db     241097 AGGCAGGAGAATCACTTGAACCTAGGAGACAGAG---GTTACAGTGAGCTGAGATCGCGC 241041

Qy       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCA 2503
              |||||||||||||||||||||| ||| ||||||||  | ||
Db     241040 CATTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCCATCTCA 240998


RESULT 10
ADQ97523/c
ID   ADQ97523 standard; DNA; 215974 BP.
XX
AC   ADQ97523;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD09-008, SEQ ID 500.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 500; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 215974 BP; 56601 A; 42897 C; 46516 G; 65586 T; 0 U; 4374 Other;

  Query Match             7.5%;  Score 186.8;  DB 12;  Length 215974;
  Best Local Similarity   83.3%;  Pred. No. 1.4e-32;
  Matches  250;  Conservative    0;  Mismatches   42;  Indels    8;  Gaps    3;

Qy       2211 AAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG------ 2264
             | ||||| | |||||||||  ||||||||| |||||||||||||||||||||||
Db     136836 ATAAATATCTCGGCCAGGCGTGGTGGCTCACTCCTGTAATCCCAGCACTTTGGGAGGCCA 136777

Qy       2265 AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACC 2324
             ||||||| |||||||||||||||||| ||||| |||||||||||||| |||||  ||| |
Db     136776 AGGTGGGCGGATCACCTGAGGTCGGCAGTTCAAGACCAGCCTGACCAACATGG-AGAAAC 136718

Qy       2325 CCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCC 2384
             ||| |||||||||||||||||||||||||| |  ||||| |||| ||||||||||||||
Db     136717 CCCGTCTCTACTAAAAACACAAAAAATTAGCCAGATGTGGTAGCACGTGCCTGTAATCCC 136658

Qy       2385 AGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCG 2443
             |||||||||||||| ||||| | |||||||  |||||| |||| ||| || ||  |||||
Db     136657 AGCTACTCAGGAGACTGAGACAGGAGAATCACTTGAATCCGGGAGGAGGTGGAGGTTGCA 136598

Qy       2444 GTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCA 2503
             ||||| ||| ||  | | ||||||||| |||||||||||| |||  |||||||  | ||
Db     136597 GTGAGTTGACATCACGCCATTGCACTCTAGCCTGGGCAACAAGAGCGAAACTCCGTCTCA 136538


RESULT 11
AAK85292
ID   AAK85292 standard; DNA; 2680 BP.
XX
AC   AAK85292;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40104.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
```

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 40104; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 2680 BP; 723 A; 594 C; 630 G; 733 T; 0 U; 0 Other;

Query Match 7.4%; Score 186.4; DB 4; Length 2680;

```
  Best Local Similarity   84.6%;  Pred. No. 4.3e-33;
  Matches  259;  Conservative    0;  Mismatches   36;  Indels   11;  Gaps    4;

Qy       2205 TTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 2264
              || || ||| | |    ||||  ||| ||||||||||||||||||||||||||||||||
Db       1131 TTTTAAAAAGAAAAAGCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 1190

Qy       2265 ------AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGT 2318
                    ||||||| |||||||||| ||| ||||||||||||||||||||| |||||
Db       1191 AGGCCAAGGTGGGCGGATCACCTGATGTCAGGAGTTCGAGACCAGCCTGACCAACATGG- 1249

Qy       2319 GGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGT 2378
              ||| || | ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1250 AGAAACCTCGTCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCGCATGCCTGT 1309

Qy       2379 AATCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTT 2437
              |||||||||||||| |||||||||||| ||||||| |||||||||  | |||  |||
Db       1310 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTAGGAGGCAGAG--- 1366

Qy       2438 GTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
              ||||| |||||||||||||||| | ||||| |||||||||||| | |||| |||||||
Db       1367 GTTGCAGTGAGCTGAGATTGCACCACTGCACCCCAGCCTGGGCAATAAGAGTGAAACTCC 1426

Qy       2498 CCCCCA 2503
                | ||
Db       1427 ATCTCA 1432


RESULT 12
ABA15481/c
ID   ABA15481 standard; DNA; 5788 BP.
XX
AC   ABA15481;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7812.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   20-OCT-2000; 2000US-0242221P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
```

```
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250391P.
PR   01-DEC-2000; 2000US-0251160P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
PT   and metastases.
XX
PS   Disclosure; SEQ ID NO 7812; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5788 BP; 1326 A; 1585 C; 1382 G; 1495 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 186.4;  DB 5;  Length 5788;
  Best Local Similarity   87.2%;  Pred. No. 5.6e-33;
  Matches  253;  Conservative    0;  Mismatches   26;  Indels   11;  Gaps    4;

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGG------GAGGTGGGTGG 2274
             |||||||||||  ||||||||  ||||||||||||||||||||      ||||||||||
Db       3684 CGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGGTGAGGTGGGTGG 3625

Qy       2275 ATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTA 2334
             ||||||||||||||||||||||||||||||||||||||||||| |||||  ||| |  ||||||||
Db       3624 ATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAACATGG-AGAAACGACATCTCTA 3566

Qy       2335 CTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAG 2394
             |||||||  ||||||||||||| ||||||||||||||||||||||||||||||||||||  |
Db       3565 CTAAAAATACAAAAAATTAGCTGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGG 3506

Qy       2395 GAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAG 2453
             |||||||||||  |||||||  ||||||  ||||  |||  ||||   |||||  |||| |||||
Db       3505 GAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCGGAG---GTTGCAGTGAACTGAG 3449

Qy       2454 ATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCA 2503
             || |||||||||||||||||||||||||||||||  |||  || ||||  | ||
Db       3448 ATCGCACTATTGCACTCCAGCCTGGGCAACAAGAGCGAACCTCCATCTCA 3399


RESULT 13
ABA15482/c
ID   ABA15482 standard; DNA; 5788 BP.
XX
AC   ABA15482;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7813.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
```

PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
PS Disclosure; SEQ ID NO 7813; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of

```
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5788 BP; 1326 A; 1585 C; 1382 G; 1495 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 186.4;  DB 5;  Length 5788;
  Best Local Similarity   87.2%;  Pred. No. 5.6e-33;
  Matches  253;  Conservative    0;  Mismatches   26;  Indels   11;  Gaps    4;

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGG------GAGGTGGGTGG 2274
             |||||||||| |||||||| |||||||||||||||||||||      ||||||||||
Db       3684 CGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGGTGAGGTGGGTGG 3625

Qy       2275 ATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTA 2334
             |||||||||||||||||||||||||||||||||||| |||||  ||| |  ||||||||
Db       3624 ATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAACATGG-AGAAACGACATCTCTA 3566

Qy       2335 CTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAG 2394
             ||||||| |||||||||||| ||||||||||||||||||||||||||||||||||  |
Db       3565 CTAAAAATACAAAAAATTAGCTGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGG 3506

Qy       2395 GAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAG 2453
             |||||||||||  |||||||  |||||| |||| ||| ||||   ||||| |||| |||||
Db       3505 GAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCGGAG---GTTGCAGTGAACTGAG 3449

Qy       2454 ATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCA 2503
             || ||||||||||||||||||||||||||| |||  || ||||  | ||
Db       3448 ATCGCACTATTGCACTCCAGCCTGGGCAACAAGAGCGAACCTCCATCTCA 3399


RESULT 14
AAK87584/c
ID   AAK87584 standard; DNA; 14417 BP.
XX
AC   AAK87584;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42396.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
```

```
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 42396; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 14417 BP; 3988 A; 3462 C; 3356 G; 3611 T; 0 U; 0 Other;
```

```
  Query Match             7.4%;  Score 186.4;  DB 4;  Length 14417;
  Best Local Similarity   84.6%;  Pred. No. 7.4e-33;
  Matches  259;  Conservative    0;  Mismatches   36;  Indels   11;  Gaps    4;

Qy       2205 TTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 2264
             || || ||| | |    ||||  ||| ||||||||||||||||||||||||||||||||||
Db       2710 TTTTAAAAAGAAAAAGCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 2651

Qy       2265 ------AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGT 2318
                  ||||||| |||||||||||| ||| ||||||||||||||||||||||| |||||
Db       2650 AGGCCAAGGTGGGCGGATCACCTGATGTCAGGAGTTCGAGACCAGCCTGACCAACATGG- 2592

Qy       2319 GGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGT 2378
              ||| || | ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       2591 AGAAACCTCGTCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCGCATGCCTGT 2532

Qy       2379 AATCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTT 2437
              |||||||||||||| |||||||||| ||||||| |||||||||  | |||  |||
Db       2531 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTAGGAGGCAGAG--- 2475

Qy       2438 GTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
              ||||| |||||||||||||||| | ||||| |||||||||||||| | |||| |||||||
Db       2474 GTTGCAGTGAGCTGAGATTGCACCACTGCACCCCAGCCTGGGCAATAAGAGTGAAACTCC 2415

Qy       2498 CCCCCA 2503
                  | ||
Db       2414 ATCTCA 2409
```

```
RESULT 15
AAK71816
ID   AAK71816 standard; DNA; 14417 BP.
XX
AC   AAK71816;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26628.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
```

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 26628; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 14417 BP; 3611 A; 3356 C; 3462 G; 3988 T; 0 U; 0 Other;

Query Match 7.4%; Score 186.4; DB 4; Length 14417;

Best Local Similarity 84.6%; Pred. No. 7.4e-33;
Matches 259; Conservative 0; Mismatches 36; Indels 11; Gaps 4;

```
Qy       2205 TTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 2264
              || || ||| | |   |||| ||| ||||||||||||||||||||||||||||||||||
Db      11708 TTTTAAAAAGAAAAAGCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 11767

Qy       2265 ------AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGT 2318
                    ||||||| |||||||||| ||| |||||||||||||||||||| |||||
Db      11768 AGGCCAAGGTGGGCGGATCACCTGATGTCAGGAGTTCGAGACCAGCCTGACCAACATGG- 11826

Qy       2319 GGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGT 2378
              ||| || | ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      11827 AGAAACCTCGTCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCGCATGCCTGT 11886

Qy       2379 AATCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTT 2437
              |||||||||||||| |||||||||||| ||||||| ||||||||  | |||  |||
Db      11887 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTAGGAGGCAGAG--- 11943

Qy       2438 GTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCC 2497
              ||||| |||||||||||||||| | ||||| |||||||||||| | |||| |||||||
Db      11944 GTTGCAGTGAGCTGAGATTGCACCACTGCACCCCAGCCTGGGCAATAAGAGTGAAACTCC 12003

Qy       2498 CCCCCA 2503
                | ||
Db      12004 ATCTCA 12009
```

Search completed: September 8, 2005, 23:49:50
Job time : 894.428 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 23:15:40 ; Search time 289.657 Seconds
(without alignments)
14156.413 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2494	99.5	2505	3	US-09-097-199-85	Sequence 85, Appl
	2	2424.2	96.7	2506	4	US-09-949-016-3311	Sequence 3311, Ap
	3	2424.2	96.7	6507	4	US-09-949-016-15053	Sequence 15053, A
	4	1950	77.8	2087	3	US-09-097-199-83	Sequence 83, Appl
	5	626.2	25.0	757	2	US-08-692-787-3	Sequence 3, Appli
	6	626.2	25.0	757	3	US-09-097-199-3	Sequence 3, Appli
c	7	572.2	22.8	601	4	US-09-949-016-119365	Sequence 119365,
c	8	572.2	22.8	601	4	US-09-949-016-119366	Sequence 119366,
c	9	397.6	15.9	601	4	US-09-949-016-119367	Sequence 119367,
c	10	344.2	13.7	601	4	US-09-949-016-119364	Sequence 119364,
	11	189.4	7.6	174639	4	US-09-949-016-16509	Sequence 16509, A
	12	189	7.5	601	4	US-09-949-016-121999	Sequence 121999,
	13	188	7.5	48691	4	US-09-949-016-16308	Sequence 16308, A
c	14	187.8	7.5	70947	4	US-09-949-016-15165	Sequence 15165, A
c	15	187.6	7.5	601	4	US-09-949-016-163069	Sequence 163069,
c	16	187.6	7.5	601	4	US-09-949-016-163070	Sequence 163070,
	17	187.4	7.5	601	4	US-09-949-016-122000	Sequence 122000,
c	18	186.6	7.4	27630	4	US-09-949-016-12722	Sequence 12722, A
c	19	185.4	7.4	571	4	US-09-621-976-17222	Sequence 17222, A
	20	185.4	7.4	209210	4	US-09-949-016-15094	Sequence 15094, A
	21	185	7.4	601	4	US-09-949-016-121998	Sequence 121998,
c	22	185	7.4	601	4	US-09-949-016-163071	Sequence 163071,
c	23	184.8	7.4	69062	4	US-09-949-016-13608	Sequence 13608, A
c	24	184.8	7.4	69062	4	US-09-949-016-13609	Sequence 13609, A
c	25	184.8	7.4	69687	4	US-09-949-016-12890	Sequence 12890, A
	26	184.4	7.4	601	4	US-09-949-016-63806	Sequence 63806, A
	27	184.4	7.4	601	4	US-09-949-016-63881	Sequence 63881, A
c	28	184.2	7.4	565	4	US-09-621-976-17221	Sequence 17221, A
	29	183.6	7.3	106746	4	US-09-326-402C-1	Sequence 1, Appli
	30	183.6	7.3	106746	4	US-09-326-402C-12	Sequence 12, Appl
	31	183.4	7.3	75212	4	US-09-949-016-13313	Sequence 13313, A
	32	183.4	7.3	75212	4	US-09-949-016-13314	Sequence 13314, A
	33	183.4	7.3	75212	4	US-09-949-016-13315	Sequence 13315, A
	34	183.4	7.3	87523	4	US-09-949-016-12670	Sequence 12670, A
	35	183.4	7.3	87523	4	US-09-949-016-15047	Sequence 15047, A
	36	183.4	7.3	87523	4	US-09-949-016-15048	Sequence 15048, A
	37	183.4	7.3	87523	4	US-09-949-016-15049	Sequence 15049, A
	38	183.4	7.3	87869	4	US-09-949-016-11744	Sequence 11744, A
	39	183.4	7.3	87869	4	US-09-949-016-15044	Sequence 15044, A
	40	183.4	7.3	87869	4	US-09-949-016-15045	Sequence 15045, A
	41	183.4	7.3	87869	4	US-09-949-016-15046	Sequence 15046, A
c	42	183.2	7.3	19728	4	US-09-949-016-12506	Sequence 12506, A
c	43	183.2	7.3	19728	4	US-09-949-016-15767	Sequence 15767, A
c	44	182.6	7.3	51671	4	US-09-949-016-12068	Sequence 12068, A
c	45	182.6	7.3	51671	4	US-09-949-016-15962	Sequence 15962, A

ALIGNMENTS

RESULT 1
US-09-097-199-85
; Sequence 85, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..503
US-09-097-199-85

Query Match 99.5%; Score 2494; DB 3; Length 2505;

```
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2505;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGG 1979

Qy       1981 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1980 GGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGACG 2039

Qy       2041 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2040 ACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTCT 2099

Qy       2101 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2100 AATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACGA 2159
```

```
Qy       2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2160 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACCC 2219

Qy       2221 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2220 CGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCACC 2279

Qy       2281 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2280 TGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAAA 2339

Qy       2341 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2340 ACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 2399

Qy       2401 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2400 GAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCAC 2459

Qy       2461 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
             ||||||||||||||||||||||||||||||||||||||||||||||
Db       2460 TATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2505
```

RESULT 2
US-09-949-016-3311
; Sequence 3311, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3311
; LENGTH: 2506
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3311

Query Match 96.7%; Score 2424.2; DB 4; Length 2506;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2494; Conservative 1; Mismatches 4; Indels 8; Gaps 6;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              |||||||||||||||||||||||||||||| |   |||||||||||||||||||||||
Db         61 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 118

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 178

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 238

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        239 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 298

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 358

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 418

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 478

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 538

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 598

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 658

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 718

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 778

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 838

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 898

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 958

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1018

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1078

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1138

Qy       1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1198

Qy       1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1258

Qy       1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1318

Qy       1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1378
```

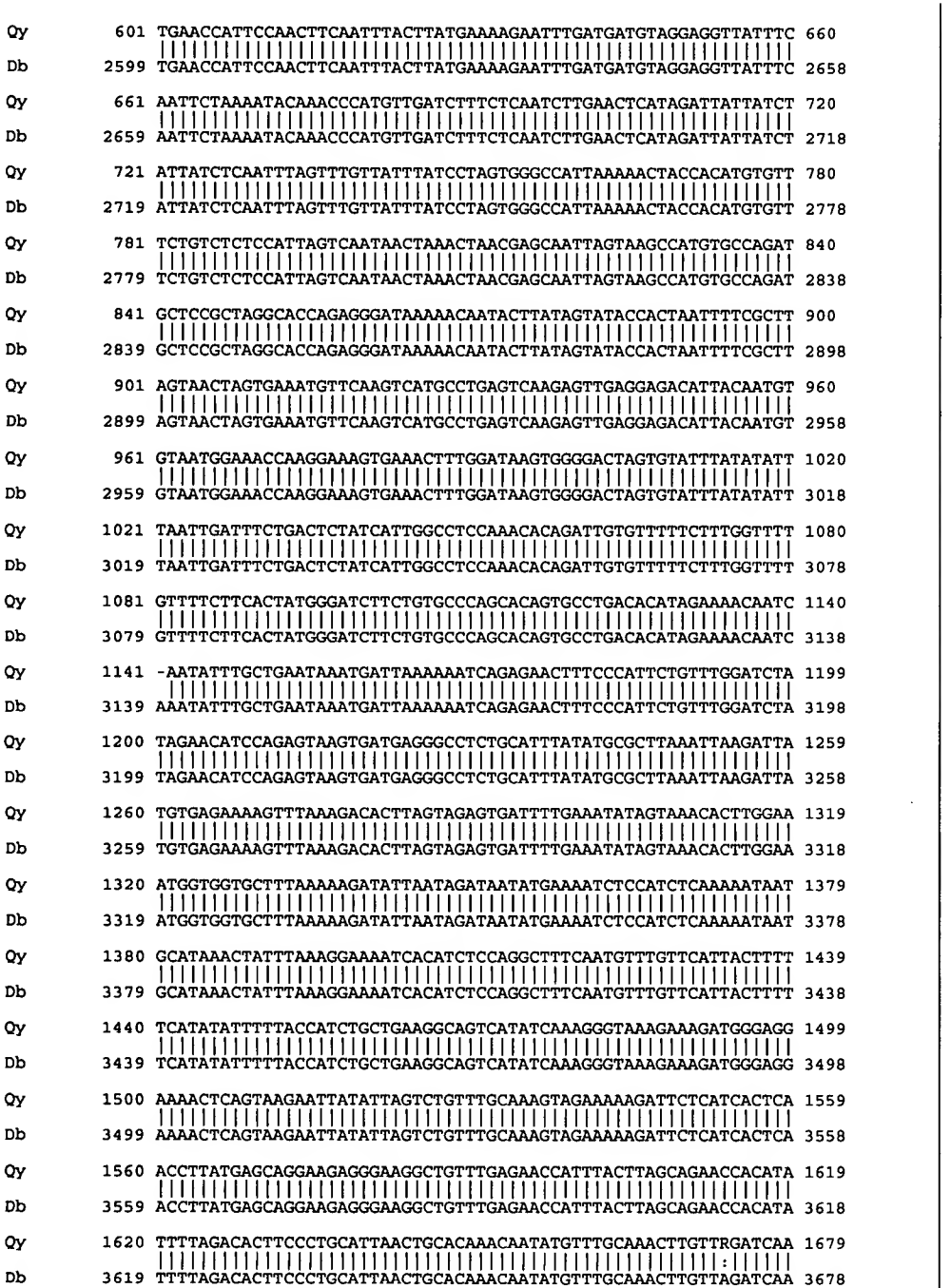
```
Qy       1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1438

Qy       1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1498

Qy       1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1558

Qy       1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1618

Qy       1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db       1619 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 1678

Qy       1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1738

Qy       1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1798

Qy       1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1858

Qy       1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1918

Qy       1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 1979
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 1978

Qy       1980 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 2039
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 2038

Qy       2040 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTC 2099
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTT-GTAATACCAGAACTC 2097

Qy       2100 TAATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 2159
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2098 TAATTGGA-CACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 2156

Qy       2160 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATATAAAAATACC 2219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2157 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATATAAAAATACC 2216

Qy       2220 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 2279
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2217 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 2276

Qy       2280 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 2339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2277 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 2336

Qy       2340 AACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 2399
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2337 AACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 2396

Qy       2400 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCA 2459
             |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db       2397 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAG---GTTGCGGTGAGCTGAGATTGCA 2453
```

```
Qy       2460 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2500
```

RESULT 3
US-09-949-016-15053
; Sequence 15053, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15053
; LENGTH: 6507
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15053

Query Match 96.7%; Score 2424.2; DB 4; Length 6507;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2494; Conservative 1; Mismatches 4; Indels 8; Gaps 6;

```
Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2001 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 2060

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
               |||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
Db       2061 -TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 2118

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2119 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 2178

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2179 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 2238

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2239 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 2298

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2299 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 2358

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2359 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 2418

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2419 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 2478

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2479 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 2538

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2539 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 2598
```



```
Qy       601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
Db      2599 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 2658

Qy       661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
Db      2659 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 2718

Qy       721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
Db      2719 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 2778

Qy       781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
Db      2779 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 2838

Qy       841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
Db      2839 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 2898

Qy       901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db      2899 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 2958

Qy       961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
Db      2959 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 3018

Qy      1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
Db      3019 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 3078

Qy      1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db      3079 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 3138

Qy      1141 -AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 1199
Db      3139 AAATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTA 3198

Qy      1200 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
Db      3199 TAGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTA 3258

Qy      1260 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 1319
Db      3259 TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAA 3318

Qy      1320 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 1379
Db      3319 ATGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT 3378

Qy      1380 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
Db      3379 GCATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 3438

Qy      1440 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 1499
Db      3439 TCATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGG 3498

Qy      1500 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 1559
Db      3499 AAAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCA 3558

Qy      1560 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 1619
Db      3559 ACCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATA 3618

Qy      1620 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAA 1679
Db      3619 TTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTAGATCAA 3678

Qy      1680 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 1739
Db      3679 CCTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTT 3738

Qy      1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
Db      3739 AACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGA 3798

Qy      1800 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 1859
Db      3799 TTTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATG 3858

Qy      1860 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 1919
Db      3859 ATCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAA 3918

Qy      1920 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 1979
Db      3919 ACTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAG 3978

Qy      1980 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 2039
Db      3979 GGGACTCAAACAGTCTTTCTTCCTTTGTCGTGTTTCTTGCTCCCAGACCAAGGCACTGAC 4038

Qy      2040 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTTGGTAATACCAGAACTC 2099
Db      4039 GACAGTACTGATACATAATTTAAAAGCACACTCCCTTCCACTTT-GTAATACCAGAACTC 4097

Qy      2100 TAATTGGACCACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 2159
Db      4098 TAATTGGA-CACCCTGAAGCTTAGGACTACCAGCCATACAAATAGTAAACTCTGTCCACG 4156

Qy      2160 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACC 2219
Db      4157 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAAAAATACC 4216

Qy      2220 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 2279
Db      4217 CCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGTGGATCAC 4276

Qy      2280 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 2339
Db      4277 CTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTCTACTAAA 4336

Qy      2340 AACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 2399
Db      4337 AACACAAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGC 4396

Qy      2400 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTGCGGTGAGCTGAGATTGCA 2459
Db      4397 TGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAG---GTTGCGGTGAGCTGAGATTGCA 4453

Qy      2460 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
Db      4454 CTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 4500
```

RESULT 4
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  83:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2087 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..503
US-09-097-199-83

  Query Match             77.8%;  Score 1950;  DB 3;  Length 2087;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1967;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
```

```
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740

Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800

Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860

Db       1800 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1859

Qy       1861 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1920

Db       1860 TCACATTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAA 1919

Qy       1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGA 1978

Db       1920 CTAACCATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCTGCACGACAACACATAAA 1977


RESULT 5
US-08-692-787-3
; Sequence 3, Application US/08692787
; Patent No. 5882864
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE
;    TITLE OF INVENTION:  DISEASE
;    NUMBER OF SEQUENCES:  82
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/692,787
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Corder, Timothy S.
;      REGISTRATION NUMBER:  38,414
;      REFERENCE/DOCKET NUMBER:  UROC:012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  757 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-692-787-3

  Query Match             25.0%;  Score 626.2;  DB 2;  Length 757;
  Best Local Similarity   94.6%;  Pred. No. 9.8e-151;
  Matches  716;  Conservative    0;  Mismatches   28;  Indels   13;  Gaps    6;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745

Db          1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805

Db         61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865

Db         121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925

Db         181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985

Db         241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036

Db         301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093

Db         361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152

Db         421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212

Db         481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272

Db         541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332

Db         601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660

Qy       2333 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392

Db         661 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429

Db         721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757


RESULT 6
US-09-097-199-3
; Sequence 3, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  757 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-097-199-3

  Query Match             25.0%;  Score 626.2;  DB 3;  Length 757;
  Best Local Similarity   94.6%;  Pred. No. 9.8e-151;
  Matches  716;  Conservative    0;  Mismatches   28;  Indels   13;  Gaps    6;

Qy       1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTAACGCT 1745
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db             1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACATTGGATTTTTCCTTAACGCT 60

Qy       1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGATTTCTC 120

Qy       1806 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           121 CCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

Qy       1866 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           181 TTCTTTCCAAAAGGAACTCTAGAAGACCAAATGCCCCGAGTTAAGAACATCAAAACTAAC 240

Qy       1926 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           241 CATCTGAAGAAACTTCCCAAGTGTAAGACTCTGCCATTAAAACATTACCGAGAGGGGACT 300

Qy       1986 CAAACAGTCTTT--CTTCCTTTGTCGTGTTTC-TTGCTCCCAGACCAAGG------CACT 2036
             |||||||||||||    | |||||||||||||| |||||||||||||| | |       |
Db           301 CAAACAGTCTTTTCTTCCCTTTGTCGTGTTTCTTTGCTCCCAGACCCAAGGCACTTGGCG 360

Qy       2037 GACGACAGTACTGATACATAATTTAAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
             |||   | |     | |   | || ||||||| ||||||||||    || ||||    |||
Db           361 GACAGTACTTGATACAATAATTTAAAAAGCACCACTCCCTTCCCACTTTGTAAATACCCA 420

Qy       2094 GAACTCTAATTGGACCACCCTGAAGCTTAGGA-CTACCAGCCATACAAATAGTAAACTCT 2152
             ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db           421 GAACTCTAATTGGACCACCCTGAAGCTTAGGACCTACCAGCCATACAAATAGTAAACTCT 480

Qy       2153 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 2212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           481 GTCCACGATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTTATATAA 540

Qy       2213 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 2272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           541 AAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGTGGGT 600

Qy       2273 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 2332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           601 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAACCCCCATCTC 660
```

```
Qy       2333 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 2392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           661 TACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTC 720

Qy       2393 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 2429
             |||||||||||||||||||||||||||||||||||||
Db           721 AGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGG 757


RESULT 7
US-09-949-016-119365/c
; Sequence 119365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119365
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119365

  Query Match             22.8%;  Score 572.2;  DB 4;  Length 601;
  Best Local Similarity   98.8%;  Pred. No. 6.2e-137;
  Matches  596;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    2;

Qy          2 ACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTAT 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA- 543

Qy         62 TGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACCA 121
              |||||||||||||||||||||||||||||| |  |  ||||||||||||||||||||||
Db        542 TGCTACAGAGCTACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACCA 484

Qy        122 GAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAG 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAG 424

Qy        182 ATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATT 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATT 364

Qy        242 CTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTT 304

Qy        302 CATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTT 361
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CAYTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTT 244

Qy        362 GACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCAC 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCAC 184

Qy        422 ATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 ATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 124

Qy        482 AACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAA 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 AACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAA 64
```

```
Qy        542 TAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACT 4

Qy        602 GAA 604
              |||
Db          3 GAA 1
```

RESULT 8
US-09-949-016-119366/c
; Sequence 119366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119366
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119366

Query Match 22.8%; Score 572.2; DB 4; Length 601;
Best Local Similarity 98.8%; Pred. No. 6.2e-137;
Matches 596; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

```
Qy         14 TCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTATTGCTACAGAGCT 73
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        601 TCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA-TGCTACAGAGCT 543

Qy         74 ACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGA 133
              |||||||||||||||||| | | ||||||||||||||||||||||||||||||||||||
Db        542 ACAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGA 484

Qy        134 TATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAGATTAAGTAACTT 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACAGATTAAGTAACTT 424

Qy        194 CCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATTCTGCTACAAGGT 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCATTCTGCTACAAGGT 364

Qy        254 ATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTTCATTGATTATTC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 ATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGTTCATTGATTATTC 304

Qy        314 CATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTTGACCATGAAAGG 373
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CAYAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACTTGACCATGAAAGG 244

Qy        374 TTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCACATTGACTTTGCA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCACATTGACTTTGCA 184

Qy        434 GTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAAAACTATACATCC 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAAAACTATACATCC 124

Qy        494 ACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAATAAAAATGGTTG 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 ACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAAATAAAAATGGTTG 64

Qy        554 AAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACTGAACCATTCCAA 613
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 AAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAACTGAACCATTCCAA 4

Qy        614 CTT 616
              |||
Db          3 CTT 1
```

RESULT 9
US-09-949-016-119367/c
; Sequence 119367, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119367
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119367

Query Match 15.9%; Score 397.6; DB 4; Length 601;
Best Local Similarity 98.6%; Pred. No. 3.6e-92;
Matches 432; Conservative 1; Mismatches 0; Indels 5; Gaps 3;

```
Qy       2069 ACTCCCTTCCACTTTGGTAATACCAGAACTCTAATTGGACCACCCTGAAGCTTAGGACTA 2128
              |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        601 ACTCCCTTCCACTTT-GTAATACCAGAACTCTAATTGGA-CACCCTGAAGCTTAGGACTA 544

Qy       2129 CCAGCCATACAAATAGTAAACTCTGTCCACGATTCACTCATCTGTGTATTTTCTATAGAT 2188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 CCAGCCATACAAATAGTAAACTCTGTCCACGATTCACTCATCTGTGTATTTTCTATAGAT 484

Qy       2189 GTTTACTAGGCGTTTGTTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTA 2248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GTTTACTAGGCGTTTGTTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTA 424

Qy       2249 ATCCCAGCACTTTGGGAGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGA 2308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ATCCCAGCACTTTGGGAGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGA 364

Qy       2309 CCAGCATGGTGGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGC 2368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCAGCATGGTGGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGC 304

Qy       2369 ACATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAG 2428
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 ACRTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAG 244

Qy       2429 GTGGAGGTTGTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAG 2488
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GTGGAG---GTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAG 187

Qy       2489 TAAAACTCCCCCCCACCC 2506
              ||||||||||||||||||
Db        186 TAAAACTCCCCCCCACCC 169
```

```
RESULT 10
US-09-949-016-119364/c
; Sequence 119364, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119364
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-119364

  Query Match             13.7%;  Score 344.2;  DB 4;  Length 601;
  Best Local Similarity   98.1%;  Pred. No. 1.7e-78;
  Matches  368;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    2;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 314

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||:||||||||||||||| |  |  ||||||||||||||||||||||||
Db        313 -TGCTACAGAGCTRCAATTCAATTTACAGTA-GTAAATATGAGGGCCTTCTTAAGGAACC 256

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 196

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 136

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 76

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 16

Qy        361 TGACCATGAAAGGTT 375
              |||||||||||||||
Db         15 TGACCATGAAAGGTT 1


RESULT 11
US-09-949-016-16509
; Sequence 16509, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16509
;   LENGTH: 174639
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(174639)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16509

  Query Match             7.6%;  Score 189.4;  DB 4;  Length 174639;
  Best Local Similarity   81.7%;  Pred. No. 1.1e-37;
  Matches  246;  Conservative    0;  Mismatches   46;  Indels    9;  Gaps    2;

Qy       2202 TTGTTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTT 2261
              ||||    |  |||      |||||||| ||||||||||||||||||||||||||||||
Db     112900 TTGTGTACTTGAAAATGGTGGGCCAGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTT 112959

Qy       2262 GGGAGGT------GGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCAT 2315
              ||||||       ||| ||||||||||||| ||||||| ||||||||||||| || |||
Db     112960 GGGAGGCTGAGGCGGGCGGATCACCTGAGGTTGGGAGTTTGAGACCAGCCTGATCAACAT 113019

Qy       2316 GGTGGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCC 2375
              || | ||||||| |||||||||||||| |||||| |||||| ||||||||||||||||||
Db     113020 GGAGAAACCCCCGTCTCTACTAAAAATACAAAAGATTAGCTGGGCGTGGTGGCACATGCC 113079

Qy       2376 TGTAATCCCAGCTACTCAGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGG 2435
              ||||||||||||||   |||||||||||| |   | | |||||||| ||||||| ||||
Db     113080 TGTAATCCCAGCTACGTGGGAGGCTGAGGCAGGAGAATCCTTGAACCGGGAAGGCGGAG- 113138

Qy       2436 TTGTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACT 2495
                ||||| |||||| |||||  | | ||||||||||||||||||| ||| |||  |||||
Db     113139 --GTTGCAGTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCTACAAGAGCGAAACT 113196

Qy       2496 C 2496
              |
Db     113197 C 113197


RESULT 12
US-09-949-016-121999
; Sequence 121999, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 121999
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-121999

  Query Match             7.5%;  Score 189;  DB 4;  Length 601;
  Best Local Similarity   84.5%;  Pred. No. 1.1e-38;
  Matches  261;  Conservative    1;  Mismatches   36;  Indels   11;  Gaps    4;

Qy       2202 TTGTTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTT 2261
```

```
                ||||||    |    ||  | |||||||||||||||||| |||||| ||||||||||||||||
Db        105 TTGTTAGGAATGTCTAAGCAGGCCAGGCACGGTGGCTGACGCCTATAATCCCAGCACTTT 164

Qy       2262 GGG------AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCAT 2315
                |||      |||||||  |||||||||||||| ||||||||||||||||||||||| |||
Db        165 GGGAGGTCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACAT 224

Qy       2316 GGTGGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCC 2375
                ||  ||| |||||| |||||||||||| |||||||||||||| ||| ||||||||||||||
Db        225 GG-AGAAACCCCAACTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCACATGCC 283

Qy       2376 TGTAATCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAG 2434
                |||||||||||||||||:|||||||||||| |||||||||||||||||||||| |||  |||
Db        284 TGTAATCCCAGCTACTCRGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGCAGAG 343

Qy       2435 GTTGTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAAC 2494
                  ||||||||||||||||||  | | |||||||||||| |||||||||  ||| |||||
Db        344 ---GTTGCGGTGAGCTGAGATGACGCCATTGCACTCCAGTCTGGGCAACCAGAGCAAAAC 400

Qy       2495 TCCCCCCCA 2503
                ||  | |||
Db        401 TCTGCACCA 409


RESULT 13
US-09-949-016-16308
; Sequence 16308, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16308
;   LENGTH: 48691
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(48691)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16308

  Query Match             7.5%;  Score 188;  DB 4;  Length 48691;
  Best Local Similarity   83.4%;  Pred. No. 1.5e-37;
  Matches  252;  Conservative    0;  Mismatches   40;  Indels   10;  Gaps    3;

Qy       2209 ATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGT 2268
                | ||||  |  | ||||  | | ||||||||||||||||||||||||||||||||||||
Db      42100 AAAAAATAAAACAGGCCGAGTATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC 42159

Qy       2269 ------GGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAA 2322
                    |||||||||||||||||| |||||| |||||||||||| ||| ||||||| ||
Db      42160 CGAGGCGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAA 42219

Qy       2323 CCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATC 2382
                |||||||||||||||||| |||||||||||||| |  |||||||  || ||||||||||
Db      42220 CCCCCATCTCTACTAAAAATACAAAAAATTAGCCAGATGTGGTGGTGCACGCCTGTAATC 42279

Qy       2383 CCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTG 2441
                |||||||||  |||||||||||| ||||||| |||||||||||| ||| ||||   ||||
Db      42280 CCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAG---GTTG 42336

Qy       2442 CGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCC 2501
                | |||||||||||| |||| | ||||||||||||||    ||| |||| |||||||  |
Db      42337 CAGTGAGCTGAGATCGCACCACTGCACTCCAGCCTGGATGACAAGAGTGAAACTCCGTCT 42396

Qy       2502 CA 2503
                ||
Db      42397 CA 42398


RESULT 14
US-09-949-016-15165/c
; Sequence 15165, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15165
;   LENGTH: 70947
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15165

  Query Match             7.5%;  Score 187.8;  DB 4;  Length 70947;
  Best Local Similarity   84.5%;  Pred. No. 1.9e-37;
  Matches  261;  Conservative    0;  Mismatches   37;  Indels   11;  Gaps    4;

Qy       2202 TTGTTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTT 2261
                ||||||    |    ||  | |||||||||||||||||| |||||| ||||||||||||||||
Db      47525 TTGTTAGGAATGTCTAAGCAGGCCAGGCACGGTGGCTGACGCCTATAATCCCAGCACTTT 47466

Qy       2262 GGG------AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCAT 2315
                |||      |||||||  |||||||||||||| ||||||||||||||||||||||| |||
Db      47465 GGGAGGTCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACAT 47406

Qy       2316 GGTGGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCC 2375
                ||  ||| ||||| |||||||||||| |||||||||||||| ||| ||||||||||||||||
Db      47405 GG-AGAAACCCCAACTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCACATGCC 47347

Qy       2376 TGTAATCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAG 2434
                |||||||||||||||||| |||||||||||| |||||||||||||||||||||| |||  |||
Db      47346 TGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGCAGAG 47287

Qy       2435 GTTGTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAAC 2494
                  ||||||||||||||||||  | | |||||||||||| |||||||||  ||| |||||
Db      47286 ---GTTGCGGTGAGCTGAGATGACGCCATTGCACTCCAGTCTGGGCAACCAGAGCAAAAC 47230

Qy       2495 TCCCCCCCA 2503
                ||  | |||
Db      47229 TCTGCACCA 47221


RESULT 15
US-09-949-016-163069/c
; Sequence 163069, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
```

```
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 163069
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-163069

  Query Match             7.5%;  Score 187.6;  DB 4;  Length 601;
  Best Local Similarity   83.1%;  Pred. No. 2.4e-38;
  Matches  251;  Conservative    1;  Mismatches   40;  Indels   10;  Gaps    3;

Qy       2209 ATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGT 2268
              | ||||  |  | ||||  | | ||||||||||||||||||||||||||||||||||||
Db        522 AAAAAATAAAACAGGCCGAGTATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC 463

Qy       2269 ------GGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGGAA 2322
                    |||||||||||||||||| |||||| |||||||||||| ||| ||||||| ||
Db        462 CGAGGCGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAA 403

Qy       2323 CCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAATC 2382
              |||||||||||||||||| |||||||||||||| |  |||||||  || ||||||||||
Db        402 CCCCCATCTCTACTAAAAATACAAAAAATTAGCCAGATGTGGTGGTGCACGCCTGTAATC 343

Qy       2383 CCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGTTG 2441
              |||||||||  |||||||||||| ||||||| |||||||||:|| ||| ||||   ||||
Db        342 CCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCYGGGAGGCGGAG---GTTG 286

Qy       2442 CGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCC 2501
              | |||||||||||| |||| | ||||||||||||||    ||| |||| |||||||   |
Db        285 CAGTGAGCTGAGATCGCACCACTGCACTCCAGCCTGGATGACAAGAGTGAAACTCCGTCT 226

Qy       2502 CA 2503
              ||
Db        225 CA 224


Search completed: September  9, 2005, 06:29:07
Job time : 378.657 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 9, 2005, 03:30:51 ; Search time 1331.01 Seconds
(without alignments)
12364.714 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2494	99.5	2505	10	US-09-974-546-85	Sequence 85, Appl
	2	1950	77.8	2087	10	US-09-974-546-83	Sequence 83, Appl
	3	626.2	25.0	757	10	US-09-974-546-3	Sequence 3, Appli
c	4	190.6	7.6	299598	19	US-10-322-696-16	Sequence 16, Appl
c	5	189.4	7.6	129447	20	US-10-719-993-6897	Sequence 6897, Ap
	6	186.4	7.4	14417	9	US-09-860-670-251	Sequence 251, App
	7	186.4	7.4	14417	17	US-10-227-646-251	Sequence 251, App
	8	186.4	7.4	14426	9	US-09-860-670-249	Sequence 249, App
	9	186.4	7.4	14426	17	US-10-227-646-249	Sequence 249, App
c	10	186.2	7.4	532	13	US-10-027-632-187685	Sequence 187685,
c	11	186.2	7.4	532	13	US-10-027-632-187686	Sequence 187686,
c	12	186.2	7.4	532	17	US-10-027-632-187685	Sequence 187685,
c	13	186.2	7.4	532	17	US-10-027-632-187686	Sequence 187686,
c	14	186.2	7.4	43623	21	US-10-741-600-17818	Sequence 17818, A
	15	184.6	7.4	450	13	US-10-027-632-266585	Sequence 266585,
	16	184.6	7.4	450	17	US-10-027-632-266585	Sequence 266585,
	17	184.4	7.4	600	21	US-10-956-157-7352	Sequence 7352, Ap
	18	184.4	7.4	1449	21	US-10-887-553A-909	Sequence 909, App
	19	184.4	7.4	1449	21	US-10-956-157-2117	Sequence 2117, Ap
	20	184.4	7.4	1599	14	US-10-153-668-439	Sequence 439, App
c	21	184.4	7.4	3079	13	US-10-027-632-112690	Sequence 112690,
c	22	184.4	7.4	3079	13	US-10-027-632-112691	Sequence 112691,
c	23	184.4	7.4	3079	17	US-10-027-632-112690	Sequence 112690,
c	24	184.4	7.4	3079	17	US-10-027-632-112691	Sequence 112691,
	25	183.8	7.3	44348	19	US-10-301-832-11	Sequence 11, Appl
	26	183.6	7.3	106746	20	US-10-856-888-1	Sequence 1, Appli
	27	183.6	7.3	106746	20	US-10-856-888-12	Sequence 12, Appl
c	28	183.6	7.3	108316	17	US-10-292-798-1789	Sequence 1789, Ap
c	29	183.6	7.3	108317	15	US-10-017-161-2143	Sequence 2143, Ap
c	30	183.4	7.3	585	13	US-10-027-632-106412	Sequence 106412,
c	31	183.4	7.3	585	17	US-10-027-632-106412	Sequence 106412,
	32	183.4	7.3	723	13	US-10-027-632-22469	Sequence 22469, A
	33	183.4	7.3	723	17	US-10-027-632-22469	Sequence 22469, A
	34	183.2	7.3	110665	20	US-10-723-860-2002	Sequence 2002, Ap
	35	183.2	7.3	110665	22	US-10-756-149-1888	Sequence 1888, Ap
	36	183.2	7.3	227968	20	US-10-723-860-1357	Sequence 1357, Ap
c	37	183	7.3	73723	19	US-10-322-281-118	Sequence 118, App
	38	182.8	7.3	463	10	US-09-918-995-23846	Sequence 23846, A
	39	182.6	7.3	89625	13	US-10-087-192-2020	Sequence 2020, Ap
	40	182.6	7.3	90614	19	US-10-331-053-64	Sequence 64, Appl
c	41	182.2	7.3	71048	13	US-10-087-192-352	Sequence 352, App
	42	182	7.3	16156	15	US-10-017-161-981	Sequence 981, App
c	43	181.8	7.3	91760	13	US-10-087-192-844	Sequence 844, App
	44	181.6	7.2	14426	9	US-09-860-670-252	Sequence 252, App
	45	181.6	7.2	14426	17	US-10-227-646-252	Sequence 252, App

ALIGNMENTS

RESULT 1
US-09-974-546-85
; Sequence 85, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199

```
;             FILING DATE: 1998-06-12
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;   INFORMATION FOR SEQ ID NO: 85:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2505 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  99..503
;        SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-974-546-85

  Query Match             99.5%;  Score 2494;  DB 10;  Length 2505;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2505;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATAATAATTTACAAAATTATTCTTCTA 60

Qy         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCTACAGAGCTACAATTCAATTTACAGTAGGCCACCATGAGGGCCTTCTTAAGGAACC 120

Qy        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180

Qy        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCTCCAGAAACTGTGCTCTTACCAT 240

Qy        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAAAGAAGTCAAAAGGCAACAGAGT 300

Qy        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCATTGATTATTCCATAGAACAGTCACACCATGCAATTCTCACACCCTTGCAGACACACT 360

Qy        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTATCTTCAGAAGCCATATTATTCA 420

Qy        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCA 480

Qy        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAAGCCTTTCCAGAAAA 540

Qy        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAGCCCAGCCAAGTAAC 600

Qy        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGAACCATTCCAACTTCAATTTACTTATGAAAAGAATTTGATGATGTAGGAGGTTATTTC 660

Qy        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTATTATCT 720

Qy        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTATCTCAATTTAGTTTGTTATTTATCCTAGTGGGCCATTAAAAACTACCACATGTGTT 780

Qy        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCTGTCTCTCCATTAGTCAATAACTAAACTAACGAGCAATTAGTAAGCCATGTGCCAGAT 840

Qy        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTCCGCTAGGCACCAGAGGGATAAAAACAATACTTATAGTATACCACTAATTTTCGCTT 900

Qy        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTAACTAGTGAAATGTTCAAGTCATGCCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960

Qy        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTAATGGAAACCAAGGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATATT 1020

Qy       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TAATTGATTTCTGACTCTATCATTGGCCTCCAAACACAGATTGTGTTTTTCTTTGGTTTT 1080

Qy       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTTTCTTCACTATGGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAACAATC 1140

Qy       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATATTTGCTGAATAAATGATTAAAAAATCAGAGAACTTTCCCATTCTGTTTGGATCTAT 1200

Qy       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGAACATCCAGAGTAAGTGATGAGGGCCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAAATATAGTAAACACTTGGAAA 1320

Qy       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGTGGTGCTTTAAAAAGATATTAATAGATAATATGAAAATCTCCATCTCAAAAATAATG 1380

Qy       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAAACTATTTAAAGGAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATATATTTTTACCATCTGCTGAAGGCAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA 1500

Qy       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACTCAGTAAGAATTATATTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA 1560

Qy       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CCTTATGAGCAGGAAGAGGGAAGGCTGTTTGAGAACCATTTACTTAGCAGAACCACATAT 1620

Qy       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db       1621 TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAAACTTGTT-GATCAAC 1679

Qy       1681 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 CTCCAACAACGACACATTCAGGAGTTAAATATTTTTCATCAAACATTGGATTTTTCCTTA 1739

Qy       1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAAGAGAT 1799

Qy       1801 TTCTCCCTGTTATAAGCAGCAAGACAAATTAGCCATTTCACTCTCAAACTTCACTAATGA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:58:29 ; Search time 5412.94 Seconds
(without alignments)
17622.418 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gaccttaaatatatcgaggt.....agtaaaactcccccccaccc 2506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	198.8	7.9	700	9	AG163691	AG163691 Pan trogl
c	2	194.2	7.7	490	8	AQ416174	AQ416174 RPCI-11-1
	3	187.4	7.5	586	6	CB217256	NISC_nb01
	4	186.4	7.4	487	7	CN275741	170006000
c	5	185.8	7.4	754	7	CN264392	170004243
c	6	185.8	7.4	818	6	CD104583	AGENCOURT
	7	185.2	7.4	390	7	CN272500	170006000
c	8	184.8	7.4	402	5	BQ631269	il18h07.x
c	9	184.8	7.4	472	5	BU790005	in49d06.x
c	10	184.4	7.4	344	1	AI590404	ts10b03.x
	11	184.4	7.4	541	5	BM993720	BM993720 UI-H-DH0-
c	12	184.4	7.4	613	5	BM990533	UI-H-DH0-
c	13	184.4	7.4	733	6	CA308850	CA308850 UI-H-FT1-
	14	184.4	7.4	902	2	BF034649	BF034649 601455087
	15	184.4	7.4	906	5	BQ227117	BQ227117 AGENCOURT
c	16	184.4	7.4	949	5	BQ950110	BQ950110 AGENCOURT
	17	183.8	7.3	918	2	BF526838	BF526838 602070394
c	18	183.2	7.3	686	9	AG091944	AG091944 Pan trogl
c	19	183.2	7.3	2900	3	HSM804005	AL832694 Homo sapi
	20	182.6	7.3	521	5	BQ021987	BQ021987 UI-1-BBip
	21	182.2	7.3	969	3	CR623410	full-leng
	22	182	7.3	342	1	AJ573255	AJ573255 AJ573255
c	23	181.6	7.2	357	1	AI031759	AI031759 ow39d11.x
c	24	181.6	7.2	366	5	BU736145	BU736145 UI-E-DX0-
c	25	181.4	7.2	826	5	BQ220083	BQ220083 AGENCOURT
	26	181.2	7.2	654	9	AG051672	AG051672 Pan trogl
	27	181	7.2	627	1	AV763460	AV763460 AV763460
	28	181	7.2	628	6	CA306131	CA306131 UI-H-FT1-
c	29	181	7.2	668	7	CK003906	CK003906 AGENCOURT
	30	181	7.2	716	5	BM969125	BM969125 UI-CF-EN0
	31	180.8	7.2	681	5	BQ016777	BQ016777 UI-H-DI0-
	32	180.8	7.2	862	4	BG762519	BG762519 602733970
	33	180.8	7.2	876	6	CD556766	CD556766 AGENCOURT
	34	180.8	7.2	909	4	BG034698	BG034698 602300050
	35	180.8	7.2	3128	3	CR627380	CR627380 Homo sapi
	36	180.8	7.2	3243	3	HSM805448	AL834376 Homo sapi
c	37	180.6	7.2	346	5	BM991496	BM991496 UI-H-DH0-
c	38	180.6	7.2	582	5	BP250579	BP250579 BP250579
c	39	180	7.2	392	1	AI311796	AI311796 qo91f05.x
c	40	180	7.2	887	5	BQ423685	BQ423685 AGENCOURT
	41	180	7.2	941	5	BX451504	BX451504 BX451504
	42	179.8	7.2	487	6	CD106585	CD106585 AGENCOURT
	43	179.8	7.2	678	6	CB270171	CB270171 1009078 H
	44	179.8	7.2	679	2	BE392753	BE392753 601307819
c	45	179.8	7.2	694	6	CA420398	CA420398 UI-H-FG0-

ALIGNMENTS

RESULT 1
AG163691/c
LOCUS AG163691 700 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-030I03.TJ, genomic survey sequence.
ACCESSION AG163691
VERSION AG163691.1 GI:16693369
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 700)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .700
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-030I03.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

Query Match 7.9%; Score 198.8; DB 9; Length 700;

```
  Best Local Similarity   87.5%;  Pred. No. 2e-26;
  Matches  266;  Conservative    0;  Mismatches   27;  Indels   11;  Gaps    4;

Qy       2207 ATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG-- 2264
              | || |||| ||| | ||| |||| ||||||||||||||||||||||||||||||||||
Db        423 ACATTAAAAAACCGCAGCCGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG 364

Qy       2265 ----AGGTGGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATGGTGG 2320
                  |||||||||||||||||||||||| |||||||||||||||||||||||| |||||  |
Db        363 GCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGG-AG 305

Qy       2321 AACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCTGTAA 2380
              || |||||||||||||||||| |||||||| ||||||| ||||||||||||||||||||
Db        304 AAACCCCATCTCTACTAAAAATACAAAAAAGTAGCCGGACGTGGTGGCACATGCCTGTAA 245

Qy       2381 TCCCAGCTACTCAGGAGGCTGAGGC-GGAGAATTGCTTGAACCCGGAAGGTGGAGGTTGT 2439
              |||||||||||||||||||||||| |||||||||||||||||| | |||  ||||||||
Db        244 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCAGAGGTTGT 185

Qy       2440 TGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCC 2499
                |||||||||||||  | | |||||||||||||||||||||||| ||| |||||||  |
Db        184 ---GGTGAGCTGAGATGACGCCATTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCTGC 128

Qy       2500 CCCA 2503
              | ||
Db        127 CTCA 124


RESULT 2
AQ416174/c
LOCUS       AQ416174                490 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-195E6.TV RPCI-11 Homo sapiens genomic clone RPCI-11-195E6,
            genomic survey sequence.
ACCESSION   AQ416174
VERSION     AQ416174.1  GI:4470298
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-195E6.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7574597"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-195E6"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             7.7%;  Score 194.2;  DB 8;  Length 490;
  Best Local Similarity   84.0%;  Pred. No. 1.5e-25;
  Matches  258;  Conservative    0;  Mismatches   38;  Indels   11;  Gaps    3;

Qy       2203 TGTTATATAAAAATACCCCGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTG 2262
              | || || ||| ||  || |||||| |||||||||||| ||||||||||||||||||||
Db        348 TATTTTAAAAATATTTCCAGGCCAGACACGGTGGCTCACACCTGTAATCCCAGCACTTTG 289

Qy       2263 GGAGGT------GGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAGCATG 2316
              ||||||      |||||||||||||||||||||||||||||||||||| ||| ||||
Db        288 GGAGGTTGAGGCAGGTGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGGCCAACATG 229

Qy       2317 GTGGAACCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGGCGTGGTGGCACATGCCT 2376
              ||| ||||||  |||||||||||| |||||||||||| |||| |||||||   ||||
Db        228 GTGAAACCCCATTCTCTACTAAAAATACAAAAAATTAGCTGGGCATGGTGGCAGGCGCCT 169

Qy       2377 GTAATCCCAGCTACTCAGGAGGCTGAGGCGGAGAATTGCTTGAACCCGGAAGGTGGAGGT 2436
              |||||||||||||| ||||||||||||  ||||| ||||||||| || ||||| ||
Db        168 ATAATCCCAGCTACTCGGGAGGCTGAGGC--AGAATAGCTTGAACCTGGGAGGTGAAG-- 113

Qy       2437 TGTTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTC 2496
              |||||  |||||||||||| || | | |||||||||||||||||||| |||  ||||||
Db        112 -GTTGCAGTGAGCTGAGATCGCCCCACTGCACTCCAGCCTGGGCAACAAGAGCGAAACTC 54

Qy       2497 CCCCCCA 2503
              |  | ||
Db         53 CATCTCA 47


RESULT 3
CB217256
LOCUS       CB217256                586 bp    mRNA    linear   EST 06-FEB-2003
DEFINITION  NISC_nb01f07.x1 COGENE 6E MAN Homo sapiens cDNA clone IMAGE:5795076
            3', mRNA sequence.
ACCESSION   CB217256
VERSION     CB217256.1  GI:28265448
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM12897  row: K  column: 13
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5795076"
                     /tissue_type="mandible, pooled"
                     /dev_stage="embryo, 6 weeks postconception"
                     /lab_host="DH10B"
                     /clone_lib="COGENE 6E MAN"
                     /note="Vector: pAMP1; cDNA primed using oligo-dT primer,
                     directionally cloned into UDG sites of pAMP1. Size
                     selected for insert sizes ranging from 0.2-2.0 kb.
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:07:44 ; Search time 2414 Seconds
(without alignments)
2709.799 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174502_20945/app_query.fasta_1.654
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_2647_@runat_07092005_174502_20945 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	2087	6	AR146835	AR146835 Sequence
2	694	100.0	2103	9	AF189270	AF189270 Homo sapi
3	694	100.0	2496	9	BC069109	BC069109 Homo sapi
4	694	100.0	2505	6	AR146836	AR146836 Sequence
5	694	100.0	2512	9	AF189269	AF189269 Homo sapi
6	694	100.0	129227	9	HS171N11	AL031433 Human DNA
7	694	100.0	186698	2	AC036236	AC036236 Homo sapi
c 8	99.5	14.3	259920	2	AC096345	AC096345 Rattus no
9	95	13.7	9833	9	AF291597	AF291597 Homo sapi
10	94	13.5	9432	9	AF291608	AF291608 Homo sapi
11	94	13.5	9817	9	AF291631	AF291631 Homo sapi
12	94	13.5	9817	9	AF291632	AF291632 Homo sapi
13	94	13.5	9830	9	AF291644	AF291644 Homo sapi
14	94	13.5	9833	9	AF291587	AF291587 Homo sapi
15	94	13.5	9833	9	AF291588	AF291588 Homo sapi
16	94	13.5	9833	9	AF291589	AF291589 Homo sapi
17	94	13.5	9833	9	AF291590	AF291590 Homo sapi
18	94	13.5	9833	9	AF291591	AF291591 Homo sapi
19	94	13.5	9833	9	AF291592	AF291592 Homo sapi
20	94	13.5	9833	9	AF291593	AF291593 Homo sapi
21	94	13.5	9833	9	AF291594	AF291594 Homo sapi
22	94	13.5	9833	9	AF291595	AF291595 Homo sapi
23	94	13.5	9833	9	AF291598	AF291598 Homo sapi
24	94	13.5	9833	9	AF291599	AF291599 Homo sapi
25	94	13.5	9833	9	AF291600	AF291600 Homo sapi
26	94	13.5	9833	9	AF291601	AF291601 Homo sapi
27	94	13.5	9833	9	AF291602	AF291602 Homo sapi
28	94	13.5	9833	9	AF291603	AF291603 Homo sapi
29	94	13.5	9833	9	AF291604	AF291604 Homo sapi
30	94	13.5	9833	9	AF291605	AF291605 Homo sapi
31	94	13.5	9833	9	AF291606	AF291606 Homo sapi
32	94	13.5	9833	9	AF291607	AF291607 Homo sapi
33	94	13.5	9833	9	AF291609	AF291609 Homo sapi
34	94	13.5	9833	9	AF291610	AF291610 Homo sapi
35	94	13.5	9833	9	AF291612	AF291612 Homo sapi
36	94	13.5	9833	9	AF291614	AF291614 Homo sapi
37	94	13.5	9833	9	AF291615	AF291615 Homo sapi
38	94	13.5	9833	9	AF291618	AF291618 Homo sapi
39	94	13.5	9833	9	AF291620	AF291620 Homo sapi
40	94	13.5	9833	9	AF291621	AF291621 Homo sapi
41	94	13.5	9833	9	AF291623	AF291623 Homo sapi
42	94	13.5	9833	9	AF291624	AF291624 Homo sapi
43	94	13.5	9833	9	AF291625	AF291625 Homo sapi
44	94	13.5	9833	9	AF291626	AF291626 Homo sapi
45	94	13.5	9833	9	AF291627	AF291627 Homo sapi

ALIGNMENTS

RESULT 1
AR146835
LOCUS AR146835 2087 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 83 from patent US 6218529.
ACCESSION AR146835
VERSION AR146835.1 GI:15110024
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2087)
AUTHORS An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
TITLE Biomarkers and targets for diagnosis, prognosis and management of prostate, breast and bladder cancer
JOURNAL Patent: US 6218529-A 83 17-APR-2001;
FEATURES Location/Qualifiers
source 1. .2087
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Alignment Scores:

Pred. No.:	2.18e-67	Length:	2087
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

```
US-09-974-546C-86 (1-135) x AR146835 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
AF189270
LOCUS       AF189270                2103 bp    mRNA    linear   PRI 22-MAR-2001
DEFINITION  Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION   AF189270
VERSION     AF189270.1  GI:10441603
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2103)
  AUTHORS   An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
            Wright,G.L. Jr. and Veltri,R.W.
  TITLE     Cloning and characterization of UROC28, a novel gene overexpressed
            in prostate, breast, and bladder cancers
  JOURNAL   Cancer Res. 60 (24), 7014-7020 (2000)
  MEDLINE   21028101
   PUBMED   11156405
REFERENCE   2  (bases 1 to 2103)
  AUTHORS   An,G. and Veltri,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
            Parkway, Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1. .2103
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q23-q24"
                     /tissue_type="prostate cancer"
     gene            1. .2103
                     /gene="UC28"
     CDS             99. .506
                     /gene="UC28"
                     /codon_start=1
                     /product="UC28 protein"
                     /protein_id="AAG17118.1"
                     /db_xref="GI:10441604"
                     /translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNPPRLPGILAPETV
                     LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     misc_feature    1961. .2103
                     /gene="UC28"
                     /note="alternative splicing"
     polyA_signal    2047. .2052
                     /gene="UC28"
ORIGIN

Alignment Scores:
Pred. No.:                2.2e-67        Length:          2103
Score:                    694.00         Matches:         135
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-09-974-546C-86 (1-135) x AF189270 (1-2103)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 3
BC069109
LOCUS       BC069109                2496 bp    mRNA    linear   PRI 23-JUN-2004
DEFINITION  Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cDNA
            clone IMAGE:7216926).
ACCESSION   BC069109
VERSION     BC069109.1  GI:46575743
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2496)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
```

```
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2496)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Anup Madan, University of Iowa
            cDNA Library Preparation: Anup Madan, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Neurogenomics Research Lab,
            200 B EMRB, University of Iowa, Iowa City, IA-52242
            anup-madan@uiowa.edu
            Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0.
FEATURES              Location/Qualifiers
     source           1. .2496
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:7216926"
                      /tissue_type="Synthetic constructs"
                      /clone_lib="NIH_MGC_242"
                      /lab_host="XL10"
                      /note="Vector: pUC19"
ORIGIN

Alignment Scores:
Pred. No.:                2.66e-67        Length:         2496
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-974-546C-86 (1-135) x BC069109 (1-2496)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 155

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 215

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 275

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 335

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 395

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        396 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 455

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        456 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 500

RESULT 4
AR146836
LOCUS       AR146836                2505 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 85 from patent US 6218529.
ACCESSION   AR146836
VERSION     AR146836.1  GI:15110025
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2505)
  AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
  TITLE     Biomarkers and targets for diagnosis, prognosis and management of
            prostate, breast and bladder cancer
  JOURNAL   Patent: US 6218529-A 85 17-APR-2001;
FEATURES              Location/Qualifiers
     source           1. .2505
                      /organism="unknown"
                      /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.68e-67        Length:         2505
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-974-546C-86 (1-135) x AR146836 (1-2505)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 5
AF189269
LOCUS       AF189269                2512 bp    mRNA    linear   PRI 22-MAR-2001
DEFINITION  Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION   AF189269
VERSION     AF189269.1  GI:10441601
KEYWORDS    .
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2512)
  AUTHORS   An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
            Wright,G.L. Jr. and Veltri,R.W.
  TITLE     Cloning and characterization of UROC28, a novel gene overexpressed
            in prostate, breast, and bladder cancers
  JOURNAL   Cancer Res. 60 (24), 7014-7020 (2000)
  MEDLINE   21028101
   PUBMED   11156405
REFERENCE   2  (bases 1 to 2512)
  AUTHORS   An,G. and Veltri,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
            Parkway, Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1. .2512
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q23-q24"
                     /tissue_type="prostate cancer"
     gene            1. .2512
                     /gene="UC28"
     CDS             99. .506
                     /gene="UC28"
                     /codon_start=1
                     /product="UC28 protein"
                     /protein_id="AAG17117.1"
                     /db_xref="GI:10441602"
                     /translation="MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETV
                     LLPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSS
                     EAILFTLTLQLTQTLGLECCLLYLSKTIHPQII"
     polyA_signal    2487. .2492
                     /gene="UC28"
ORIGIN

Alignment Scores:
Pred. No.:              2.68e-67        Length:         2512
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-974-546C-86 (1-135) x AF189269 (1-2512)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458
```

```
Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 6
HS171N11
LOCUS       HS171N11              129227 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
            complete sequence.
ACCESSION   AL031433
VERSION     AL031433.4  GI:4826442
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129227)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 13, 1999 this sequence version replaced gi:4582115.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-171N11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .129227
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q23"
                     /clone="RP1-171N11"
                     /clone_lib="RPCI-1"
ORIGIN

Alignment Scores:
Pred. No.:              2.19e-65        Length:         129227
Score:                  694.00          Matches:        135
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
```

```
DB:                 9             Gaps:          0

US-09-974-546C-86 (1-135) x HS171N11 (1-129227)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29892 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 29951

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29952 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 30011

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30012 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 30071

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30072 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 30131

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30132 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 30191

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30192 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 30251

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db      30252 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 30296

RESULT 7
AC036236
LOCUS       AC036236               186698 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-133O15, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC036236
VERSION     AC036236.2  GI:9958166
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186698)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:7523932.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0133O15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183373 bases at least Q40
            Consensus quality: 184432 bases at least Q30
            Consensus quality: 185064 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 186098; sum-of-contigs
            Quality coverage: 5.84 in Q20 bases; agarose-fp
            Quality coverage: 5.57 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4769: contig of 4769 bp in length
            *     4770     4869: gap of unknown length
            *     4870    11657: contig of 6788 bp in length
            *    11658    11757: gap of unknown length
            *    11758    28189: contig of 16432 bp in length
            *    28190    28289: gap of unknown length
            *    28290    53341: contig of 25052 bp in length
            *    53342    53441: gap of unknown length
            *    53442   102327: contig of 48886 bp in length
            *   102328   102427: gap of unknown length
            *   102428   182781: contig of 80354 bp in length
            *   182782   182881: gap of unknown length
            *   182882   186698: contig of 3817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186698
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-133O15"
     misc_feature    1. .4769
                     /note="assembly_name:Contig10"
     misc_feature    4870. .11657
                     /note="assembly_name:Contig11"
     misc_feature    11758. .28189
                     /note="assembly_name:Contig12"
     misc_feature    28290. .53341
                     /note="assembly_name:Contig13"
     misc_feature    53442. .102327
                     /note="assembly_name:Contig14"
     misc_feature    102428. .182781
                     /note="assembly_name:Contig15
                     clone_end:SP6
                     vector_side:right"
     misc_feature    182882. .186698
                     /note="assembly_name:Contig9"
ORIGIN

Alignment Scores:
Pred. No.:              3.3e-65        Length:         186698
Score:                  694.00         Matches:        135
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     2              Gaps:           0

US-09-974-546C-86 (1-135) x AC036236 (1-186698)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159252 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 159311

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159312 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 159371

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159372 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 159431

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     159432 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 159491

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159492 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 159551

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159552 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 159611

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db     159612 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 159656
```

RESULT 8
AC096345/c

```
LOCUS       AC096345               259920 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-18J15, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC096345
VERSION     AC096345.6  GI:30521870
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 259920)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 259920)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 259920)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24818617.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEYU
                Center clone name: CH230-18J15
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 240867 bases at least Q40
                Consensus quality: 244203 bases at least Q30
                Consensus quality: 245971 bases at least Q20
                Estimated insert size: 252807; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   258395: contig of 258395 bp in length
            *    258396   258495: gap of unknown length
            *    258496   259920: contig of 1425 bp in length.
FEATURES             Location/Qualifiers
     source          1. .259920
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-18J15"
     misc_feature    1. .1739
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    8054. .8862
                     /note="clone_boundary
                     clone_end:T7
```

```
                    site:EcoRI
                    end_sequence:BH359975"
     misc_feature   complement(248170. .249015)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH360032"
     misc_feature   249451. .251486
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   256461. .258395
                    /note="wgs_end_extension
                    clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              98.4          Length:         259920
Score:                  99.50         Matches:        38
Percent Similarity:     40.54%        Conservative:   22
Best Local Similarity:  25.68%        Mismatches:     41
Query Match:            14.34%        Indels:         47
DB:                     2             Gaps:           8

US-09-974-546C-86 (1-135) x AC096345 (1-259920)

Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
              :::||||||:::          |||:::     |||:::|||     ||||||   |||
Db     195470 TTTGAGGATGTGCAGAGGACGATATACTTTTCACAACTAAGGTTAGAAAGACATCATCTG 195411

Qy         30 SerAsnPheProArgLeuProGlyIle-----------LeuAlaProGluThrValLeu 45
              |||      |||      |||   :::                  |||:::   ::::::
Db     195410 TCCACAATGCCACATTTTCCAGAAGTAGGAAGACTGACATTTCAGCCAAAGAGAATTATT 195351

Qy         46 --------------LeuProPheCysTyr---------------------------- 50
                           |||||||||   |||
Db     195350 TCCAAATTCAGTTTCCTCCCTTTCCTGTATCCAGCTCACTAATTCCCACATATTTCTTCT 195291

Qy         51 -----------------LysValPheArgLysLys-----------GluLysValLys 60
                             :::|||   |||:::                |||:::
Db     195290 TTCCTGCCTTCTGTTTCAAATCTCTTTTGGAAAAGAACTGAGTAGGTGAATAAAATATCA 195231

Qy         61 ArgSerGln------LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHis 78
              |||:::|||      :::          :::   |||      :::   |||
Db     195230 AGAGCCCAGTGCTGGCGACAACAGAGATCACTCAGCTACTGCACCCAAATATCTCCCTCT 195171

Qy         79 AlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
                 :::||||||||||          ::::::||||||   :::          |||
Db     195170 TTCCTCCTTACACCAGCT---------ATTTCCATGAAGTCTACA---------GTATCT 195129

Qy         99 SerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGly 118
              :::   ||||||   |||:::                              |||
Db     195128 GCTGGTTCAAGCTCTGCACTT-----------------------------ACGTACAAT 195099

Qy        119 LeuGluCysCysLeuLeuTyrLeu 126
              |||:::   ||||||   ||||||
Db     195098 TTACAGAAGTGCTTAGCATATCTT 195075

RESULT 9
AF291597
LOCUS       AF291597               9833 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Japanese 18 22q11.2 noncoding genomic
            sequence.
ACCESSION   AF291597
VERSION     AF291597.1  GI:10953199
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Japanese 18"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9833
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              8.07          Length:         9833
Score:                  95.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.69%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-86 (1-135) x AF291597 (1-9833)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::            |||:::|||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGRTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                    :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||      |||          :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109
                 |||                                        |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly------------------------------ 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                         |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::      |||:::      |||      ::::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 10
AF291608
LOCUS       AF291608               9432 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Berg 26 22q11.2 noncoding genomic sequence.
```

```
ACCESSION   AF291608
VERSION     AF291608.1  GI:10953210
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9432)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9432)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9432
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Berg 26"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9432
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              9.95          Length:         9432
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-86 (1-135) x AF291608 (1-9432)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||     :::          |||:::|||||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT------------------------------------------ 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                  :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||      |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109
                |||                                         |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly------------------------------- 118
               :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
```

```
                                                                        |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy         120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
               :::      |||:::      |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 11
AF291631
LOCUS       AF291631               9817 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate South African MeJe 22q11.2 noncoding genomic
            sequence.
ACCESSION   AF291631
VERSION     AF291631.1  GI:10953233
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9817)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9817)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9817
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="South African MeJe"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9817
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:         9817
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-86 (1-135) x AF291631 (1-9817)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||     :::          |||:::|||||||||||||:::
Db       2939 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 2998

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       2999 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3034

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3035 GTAGATCTGGGATTT------------------------------------------ 3049

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                  :::|||
Db       3050 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3082

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
```

```
Db       3083 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3142

Qy        104 AlaIle----------------------------------------LeuPheThrLeu 109

Db       3143 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3202

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118

Db       3203 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3262

Qy        119 ----------------------------------------------------------Leu 119

Db       3263 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3322

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132

Db       3323 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3361
```

```
RESULT 12
AF291632
LOCUS       AF291632               9817 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate South African Kung 22q11.2 noncoding genomic
            sequence.
ACCESSION   AF291632
VERSION     AF291632.1  GI:10953234
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9817)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9817)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9817
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="South African Kung"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9817
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:         9817
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-86 (1-135) x AF291632 (1-9817)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23

Db       2939 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 2998

Qy         24 LysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAlaProGluThr 43
```

```
Db       2999 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3034

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63

Db       3035 GTAGATCTGGGATTT---------------------------------------------- 3049

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83

Db       3050 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3082

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103

Db       3083 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3142

Qy        104 AlaIle----------------------------------------LeuPheThrLeu 109

Db       3143 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3202

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly---------------------------------- 118

Db       3203 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3262

Qy        119 ----------------------------------------------------------Leu 119

Db       3263 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3322

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132

Db       3323 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3361
```

```
RESULT 13
AF291644
LOCUS       AF291644               9830 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Nigerian 9 22q11.2 noncoding genomic sequence.
ACCESSION   AF291644
VERSION     AF291644.1  GI:10953246
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9830)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9830)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9830
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Nigerian 9"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9830
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:         9830
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
```

```
Best Local Similarity:  24.28%        Mismatches:    38
Query Match:            13.54%        Indels:        76
DB:                     9             Gaps:          5

US-09-974-546C-86 (1-135) x AF291644 (1-9830)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                    |||   |||:::|||                                 :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||       :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109
                 |||                                       |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                        |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::      |||:::      |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 14
AF291587
LOCUS       AF291587               9833 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Chinese 1 22q11.2 noncoding genomic sequence.
ACCESSION   AF291587
VERSION     AF291587.1  GI:10953189
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Chinese 1"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9833
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:        9833
Score:                  94.00         Matches:       42
Percent Similarity:     34.10%        Conservative:  17
Best Local Similarity:  24.28%        Mismatches:    38
Query Match:            13.54%        Indels:        76
DB:                     9             Gaps:          5

US-09-974-546C-86 (1-135) x AF291587 (1-9833)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::           |||:::|||||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT---------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                    |||   |||:::|||                                 :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||       :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle-----------------------------------------LeuPheThrLeu 109
                 |||                                       |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                        |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::      |||:::      |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377

RESULT 15
AF291588
LOCUS       AF291588               9833 bp    DNA     linear   PRI 24-OCT-2000
DEFINITION  Homo sapiens isolate Chinese 2 22q11.2 noncoding genomic sequence.
ACCESSION   AF291588
VERSION     AF291588.1  GI:10953190
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9833)
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
  MEDLINE   20481912
   PUBMED   11005839
REFERENCE   2  (bases 1 to 9833)
```

```
  AUTHORS   Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
            Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .9833
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="Chinese 2"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
     misc_feature    1. .9833
                     /note="noncoding region"
ORIGIN

Alignment Scores:
Pred. No.:              10.4          Length:         9833
Score:                  94.00         Matches:        42
Percent Similarity:     34.10%        Conservative:   17
Best Local Similarity:  24.28%        Mismatches:     38
Query Match:            13.54%        Indels:         76
DB:                     9             Gaps:           5

US-09-974-546C-86 (1-135) x AF291588 (1-9833)

Qy          4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
              |||:::   :::||||||      :::          |||:::|||||||||||||:::
Db       2955 TTTATAATTTCTCAAAAATCCTATGAGGCAGGTATGTAAATCTATATTTTGCAGATACAG 3014

Qy         24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
              ||||||   ||||||||||||||||||||||||
Db       3015 AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT---------------------- 3050

Qy         44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
              |||   |||   |||
Db       3051 GTAGATCTGGGATTT------------------------------------------- 3065

Qy         64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
                 |||   |||:::|||                                  :::|||
Db       3066 ---GAAACCGGGTTCTTAGACACTAAG----------------------GCTAGTCCT 3098

Qy         84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
              |||   |||       |||         :::      :::||||||   :::   |||
Db       3099 TTAACTACACCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCCTGTCAGGA 3158

Qy        104 AlaIle----------------------------------------LeuPheThrLeu 109
                 |||                                        |||   :::|||
Db       3159 GATATTTTTCTCTAATAAAATAGGGATGGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218

Qy        110 ThrLeuGlnLeuThrGlnThrLeuGly-------------------------------- 118
              :::      :::   |||   :::|||
Db       3219 TCATTCCTTATGGGCCAGCAGGTTGGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3278

Qy        119 ----------------------------------------------------------Leu 119
                                                                         |||
Db       3279 AAAATGGAACCACATGGTCTGTTCCTGAATTCTCCAACAGATCCCTTTGGTCCCAACCTC 3338

Qy        120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
              :::     |||:::     |||      :::||||||
Db       3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCCT 3377


Search completed: September  9, 2005, 14:45:55
Job time : 2499 secs
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:02:29 ; Search time 309 Seconds
(without alignments)
2586.295 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174501_20885/app_query.fasta_1.654
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_468_@runat_07092005_174501_20885 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	2087	4	AAS04000	Aas04000 Biomarker
2	694	100.0	2088	3	AAZ87583	Aaz87583 Prostate
3	694	100.0	2505	4	AAS04001	Aas04001 Biomarker
4	694	100.0	2506	3	AAZ87584	Aaz87584 Prostate
5	90	13.0	1113	13	ADT45647	Adt45647 Bacterial
6	90	13.0	1146	13	ADS46591	Ads46591 Bacterial
7	89.5	12.9	4019	4	ABL27798	Abl27798 Drosophil
c 8	83.5	12.0	532	3	AAC38626	Aac38626 Arabidops
c 9	83	12.0	146547	8	ABZ80817	Abz80817 Human pho
c 10	82	11.8	24601	2	AAX13160	Aax13160 Enterococ
c 11	82	11.8	24601	6	ABS98955	Abs98955 Enterococ
c 12	81.5	11.7	133462	13	ABD32622	Abd32622 Mouse can
13	81	11.7	381	8	ABX44573	Abx44573 Bovine ES
c 14	81	11.7	110000	12	ADN46845_11	Continuation (12 o
15	81	11.7	110000	12	ADN47591_09	Continuation (10 o
c 16	81	11.7	110000	12	ADN46123_11	Continuation (12 o
17	81	11.7	110000	12	ADN47209_09	Continuation (10 o
c 18	81	11.7	110000	12	ADN46464_11	Continuation (12 o
19	81	11.7	110000	12	ADN47960_09	Continuation (10 o
c 20	80.5	11.6	1093	2	AAT95761	Aat95761 Arabidops
c 21	80.5	11.6	1093	3	AAC65293	Aac65293 Arabidops
22	80	11.5	1693	12	ADQ24476	Adq24476 Human sof
c 23	79.5	11.5	262	7	ADS31066	Ads31066 Human gen
24	79.5	11.5	585	5	ABV54205	Abv54205 Human pro
c 25	79.5	11.5	1464	3	AAC47238	Aac47238 Arabidops
26	79.5	11.5	6196	2	AAV13168	Aav13168 Complete
27	79.5	11.5	6243	2	AAV13841	Aav13841 Complete
28	79.5	11.5	6503	2	AAV13169	Aav13169 Complete
29	79.5	11.5	7379	2	AAV13176	Aav13176 Complete
30	79.5	11.5	8618	2	AAV18741	Aav18741 Complete
31	79.5	11.5	8792	2	AAV18745	Aav18745 Complete
c 32	78.5	11.3	386	8	ABX45278	Abx45278 Bovine ES
33	78.5	11.3	398	4	AAI89039	Aai89039 Human pol
34	78.5	11.3	4065	12	ADQ65017	Adq65017 Novel hum
c 35	78.5	11.3	26496	5	ABA19696	Aba19696 Human ner
c 36	78	11.2	3328	4	ABL18311	Abl18311 Drosophil
c 37	77.5	11.2	389	8	ABX39152	Abx39152 Bovine ES
38	77.5	11.2	567	10	ABX56974	Abx56974 Arabidops
c 39	77.5	11.2	2411	11	ADM02309	Adm02309 Human cDN
40	77.5	11.2	2437	12	ADP22463	Adp22463 Sea-squir
c 41	77.5	11.2	2661	4	ABL26351	Abl26351 Drosophil
42	77.5	11.2	3435	4	ABL02615	Abl02615 Drosophil
c 43	77.5	11.2	3958	4	ABL07870	Abl07870 Drosophil
c 44	77.5	11.2	4122	4	ABL07868	Abl07868 Drosophil
45	77.5	11.2	6010	10	ADF50278	Adf50278 Human PFM

ALIGNMENTS

RESULT 1
AAS04000
ID AAS04000 standard; cDNA; 2087 BP.
XX
AC AAS04000;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
XX
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 99. .506
FT /*tag= a
FT /product= "Prostate cancer marker protein"
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
XX
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.

```
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   P-PSDB; AAU02174.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Claim 2; Col 117-121; 78pp; English.
XX
CC   The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC   detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC   sequences can be used as hybridisation probes and primers that
CC   specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC   (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC   nucleic acid markers can be used to produce antibodies for the detection
CC   of prostate, breast or bladder cancer. The nucleic acids can be used as
CC   targets for therapeutic intervention in these diseases, in the
CC   identification and isolation of full-length gene sequences, including
CC   regulatory elements for gene expression, from genomic human DNA
CC   libraries, as hybridisation probes for screening genomic human DNA
CC   libraries. The kits comprising the nucleic acid sequences are useful for
CC   detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ   Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.6e-76       Length:          2087
Score:                  694.00        Matches:         135
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     4             Gaps:            0

US-09-974-546C-86 (1-135) x AAS04000 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
AAZ87583
ID   AAZ87583 standard; DNA; 2088 BP.
XX
AC   AAZ87583;
```

```
XX
DT   19-APR-2000  (first entry)
XX
DE   Prostate disease marker UC Band #28.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW   diagnosis; PCR primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
DR   P-PSDB; AAY59295.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Claim 1; Page 182-183; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              2.6e-76       Length:          2088
Score:                  694.00        Matches:         135
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     3             Gaps:            0

US-09-974-546C-86 (1-135) x AAZ87583 (1-2088)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278
```

```
Qy          61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy          81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy         101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy         121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
               |||||||||||||||||||||||||||||||||||||||||||||
Db         459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
```

RESULT 3
AAS04001
ID AAS04001 standard; cDNA; 2505 BP.
XX
AC AAS04001;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.
XX
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 99. .506
FT /*tag= a
FT /product= "Prostate cancer marker protein"
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
XX
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri R;
XX
DR WPI; 2001-289849/30.
DR P-PSDB; AAU02175.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS Claim 2; Col 121-125; 78pp; English.
XX
CC The sequence represents nucleic acid biomarker UC band 28 #3, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX

SQ Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.39e-76	Length:	2505
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-974-546C-86 (1-135) x AAS04001 (1-2505)

```
Qy           1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy          21 GlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAla 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy          41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy          61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy          81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy         101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy         121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
               |||||||||||||||||||||||||||||||||||||||||||||
Db         459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503
```

RESULT 4
AAZ87584
ID AAZ87584 standard; DNA; 2506 BP.
XX
AC AAZ87584;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 splice variant.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR P-PSDB; AAY59296.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.

```
XX
PS   Claim 1; Page 184-186; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              3.39e-76       Length:          2506
Score:                  694.00         Matches:         135
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     3              Gaps:            0

US-09-974-546C-86 (1-135) x AAZ87584 (1-2506)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 5
ADT45647
ID   ADT45647 standard; cDNA; 1113 BP.
XX
AC   ADT45647;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polynucleotide #20398.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 44085; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polynucleotide used in
CC   the scope of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1113 BP; 379 A; 204 C; 242 G; 288 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.53           Length:          1113
Score:                  90.00          Matches:         34
Percent Similarity:     44.88%         Conservative:    23
Best Local Similarity:  26.77%         Mismatches:      38
Query Match:            12.97%         Indels:          32
DB:                     13             Gaps:            7

US-09-974-546C-86 (1-135) x ADT45647 (1-1113)

Qy         13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu--------- 29
              |||||||||:::         ::::::::::|||:::       |||   :::
Db         61 ATGCACAACTTA------GCCATAAAGCTGAGAGAAAGAGGGCATGAAGTGGGAATAGTC 114
```

```
Qy          30 ---SerAsnPhePro---------------------------------------------- 33
                  :::|||   |||
Db         115 ACAAACAACAGGCCCACGGGAAAGGAAGAAGAGCTTAAGAGATATGGAATAGAGCTCATA 174

Qy          34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
               ::::::||||||||::::::|||                              :::
Db         175 AAGATCCCAGGAATTATAAGTCCT----------------------TTTTAGATGTA 210

Qy          54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle---AspTyrSer 72
                           :::|||   |||::::::      ||||||:::   |||:::
Db         211 AATTTAACCTATGGATTGAAATCCTCAGAAGAGCTCAACGAATTCTTGAAAGACTTCGAC 270

Qy          73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
               |||       ||||||||||||       |||||||||                    |||
Db         271 ATAATTCATTCTCATCATGCA---TTCACACCTCTCTCCTTAAAGGCTTTAAAAGCTGGA 327

Qy          93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                  :::|||   |||  :::   |||::::::  :::|||   |||
Db         328 AAGAATATGGAAAAGGGAACCCTGCTAACAACTCACAGCATTTCCTTTGCCCATGAATCA 387

Qy         112 GlnLeuThrGlnThrLeuGly 118
               :::|||      |||||||||
Db         388 AAGCTCTGGGATACTTTGGGG 408
```

RESULT 6
ADS46591
ID ADS46591 standard; cDNA; 1146 BP.
XX
AC ADS46591;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #1334.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 25021; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1146 BP; 388 A; 207 C; 253 G; 298 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.553	Length:	1146
Score:	90.00	Matches:	34
Percent Similarity:	44.88%	Conservative:	23
Best Local Similarity:	26.77%	Mismatches:	38
Query Match:	12.97%	Indels:	32
DB:	13	Gaps:	7

US-09-974-546C-86 (1-135) x ADS46591 (1-1146)

```
Qy          13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu--------- 29
               |||||||||:::         :::::::::|||:::      |||   :::
Db          70 ATGCACAACTTA------GCCATAAAGCTGAGAGAAAGAGGGCATGAAGTGGGAATAGTC 123

Qy          30 ---SerAsnPhePro---------------------------------------------- 33
                  :::|||   |||
Db         124 ACAAACAACAGGCCCACGGGAAAGGAAGAAGAGCTTAAGAGATATGGAATAGAGCTCATA 183

Qy          34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
               ::::::||||||||::::::|||                              :::
Db         184 AAGATCCCAGGAATTATAAGTCCT----------------------TTTTAGATGTA 219

Qy          54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle---AspTyrSer 72
                           :::|||   |||::::::      ||||||:::   |||:::
Db         220 AATTTAACCTATGGATTGAAATCCTCAGAAGAGCTCAACGAATTCTTGAAAGACTTCGAC 279

Qy          73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
               |||       ||||||||||||       |||||||||                    |||
Db         280 ATAATTCATTCTCATCATGCA---TTCACACCTCTCTCCTTAAAGGCTTTAAAAGCTGGA 336

Qy          93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                  :::|||   |||  :::   |||::::::  :::|||   |||
Db         337 AAGAATATGGAAAAGGGAACCCTGCTAACAACTCACAGCATTTCCTTTGCCCATGAATCA 396

Qy         112 GlnLeuThrGlnThrLeuGly 118
               :::|||      |||||||||
Db         397 AAGCTCTGGGATACTTTGGGG 417
```

RESULT 7
ABL27798
ID ABL27798 standard; DNA; 4019 BP.
XX
AC ABL27798;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34867.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX

```
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 34867; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4019 BP; 1158 A; 966 C; 927 G; 968 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              3.93          Length:        4019
Score:                  89.50         Matches:       35
Percent Similarity:     39.84%        Conservative:  16
Best Local Similarity:  27.34%        Mismatches:    40
Query Match:            12.90%        Indels:        37
DB:                     4             Gaps:          5

US-09-974-546C-86 (1-135) x ABL27798 (1-4019)

Qy         18 HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGly 37
              |||   |||:::   ||||||::::::||||||      ||||||   |||:::|||
Db       1389 CATAGACTAAAAGCACGAAAGGTGCAACACCGTAATTATAATTTTAGACGTGTTCCAGCG 1448

Qy         38 IleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGlu 57
                  |||   |||              ::::::     :::::::|||
Db       1449 CATAGGCAGCCCATCACA---------TCCACATCGCATCAAAGAACGAAAAGAAAAGCA 1499

Qy         58 LysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHis 77
              :::      |||      |||   |||      |||                     |||
Db       1500 CGTCAGAGCAGGGAAACTAAGTGTACTTCGAAGATA--------------------CAT 1538

Qy         78 HisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCys 97
                 |||      ||||||                                  :::|||
Db       1539 TGTACAATTTACATGCCGCTG--------------------------------AGGTGC 1565

Qy         98 SerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeu 117
              :::|||   :::            |||   :::|||:::|||||||||   :::
Db       1566 ACCTCGGAAAACCTTAATTGGTTTTTAGCCTCATTAAGCTTGCAGCTAGAAAGAATGTGG 1625

Qy        118 GlyLeuGlu--------------------------CysCys----------------- 122
              |||   |||                        ||||||
Db       1626 GGTCAAGAATAGCCATCAAATATTGCTCATACGACGTGTTGTACATGCGATTAGTGTAGG 1685

Qy        123 ---LeuLeuTyrLeuSerLysThr 129
                  :::||||||      |||
Db       1686 GAAGCGATGTATCTCCCTCTTACA 1709
```

```
RESULT 8
AAC38626/c
ID   AAC38626 standard; DNA; 532 BP.
XX
AC   AAC38626;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 21658.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic pathway;
KW   promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
```

PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	1.19	Length:	532
Score:	83.50	Matches:	39
Percent Similarity:	41.48%	Conservative:	17
Best Local Similarity:	28.89%	Mismatches:	52
Query Match:	12.03%	Indels:	27
DB:	3	Gaps:	7

US-09-974-546C-86 (1-135) x AAC38626 (1-532)

```
Qy         10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
                 |||   |||           :::   :::       |||             |||
Db        445 TACAGCGACAGATGTGTTTTTGTTGTGCTCTTCTTTATCCACTCAAGTAACCATCATCAC 386

Qy         30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
```

```
                ||| ||| ||| ::: ||| ||| |||
Db        385 TATAACTTACCCAATCTAGTTTGGGTTTGGCTTCCACACGGAACA-----------TGT 338

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe--- 68
                           |||:::|||:::||||||     |||::::::|||
Db        337 AGCCATGGGTCGTCTCTTCCAGAAGAGGTAGAGCGCTCCTTCATGGCCTCCCAGTTTGTA 278

Qy         69 ---------IleAspTyrSerIleGluGlnSerHisHisAla------IleLeuThr--- 82
                   |||   :::||||||      :::   :::|||      :::||||||
Db        277 GTTTTTGTCATAACTTTCAGCATCTCACTCGCTGCAAACGCTGCAGATGTGCTGACCGGC 218

Qy         83 ------------ProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
                      |||           ||||||:::     |||     |||
Db        217 CAACCCGTGAACCCAGCCATCATCATTCTGACCCTCCATTTCCCCAGTACCTCGTGTCTC 158

Qy         99 SerLeu---------SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGln 115
              |||           ||||||   ||||||:::||||||:::
Db        157 TCTACTCTCTCTGACTCCTCACATGCTATCATGTTCACTATG----------------- 116

Qy        116 ThrLeuGlyLeuGluCysCys---------LeuLeuTyrLeuSer 127
              :::|||      :::||||||         :::|||   ||||||
Db        115 TCTCTTGCTACACAGTGTTGTTCTGCGCTGATTCTCTGCTTATCA 71

RESULT 9
ABZ80817/c
ID   ABZ80817 standard; DNA; 146547 BP.
XX
AC   ABZ80817;
XX
DT   13-JUN-2003  (first entry)
XX
DE   Human phospholipase C gamma 1 polymorphism G329u1.
XX
KW   human; single nucleotide polymorphism; SNP; phospholipase C gamma 1;
KW   PLCG1; gene; vascular disease; plasminogen activator inhibitor type 2;
KW   PAI-2; diagnosis; atherosclerosis; coronary artery disease; ischemia;
KW   myocardial infarction; stroke; thromboembolism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(64001,c)
FT                   /*tag=  a
XX
PN   WO2003007801-A2.
XX
PD   30-JAN-2003.
XX
PF   19-JUL-2002; 2002WO-US023041.
XX
PR   20-JUL-2001; 2001US-0306941P.
PR   28-AUG-2001; 2001US-0315572P.
PR   05-OCT-2001; 2001US-0327488P.
PR   14-DEC-2001; 2001US-00017128.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J;
XX
DR   WPI; 2003-278312/27.
XX
PT   Diagnosing and prognosing vascular disease, e.g. coronary artery disease
PT   and myocardial infarction, based on the detection of polymorphisms in the
PT   phospholipase C gamma 1 and plasminogen activator inhibitor type 2 genes
PT   and polypeptides.
XX
PS   Claim 44; Page 204-245; 295pp; English.
XX
CC   The invention relates to methods, agents and apparatus for diagnosing and
CC   prognosing vascular diseases based on the detection of polymorphisms in
CC   the phospholipase C gamma 1 (PLCG1) and plasminogen activator inhibitor
CC   type 2 (PAI-2) genes and polypeptides. In a preferred claim of the
CC   invention the polymorphism is a C to T nucleotide change at position
CC   64001 of the PLCG1 gene (reference sequence GI 11345540) causing an amino
CC   acid change from Threonine to Isoleucine at position 813 of the protein.
CC   This sequence represents the region of the PLCG1 gene with the T form of
CC   the polymorphism. The methods, agents and apparatus is used for
CC   diagnosing and prognosing vascular diseases e.g. atherosclerosis,
CC   coronary artery disease (especially), myocardial infarction (especially),
CC   ischemia, stroke, peripheral vascular diseases, venous thromboembolism
CC   and pulmonary embolism
XX
SQ   Sequence 146547 BP; 39552 A; 32625 C; 33339 G; 41031 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.69e+03        Length:         146547
Score:                  83.00           Matches:        19
Percent Similarity:     65.71%          Conservative:   4
Best Local Similarity:  54.29%          Mismatches:     10
Query Match:            11.96%          Indels:         2
DB:                     8               Gaps:           1

US-09-974-546C-86 (1-135) x ABZ80817 (1-146547)

Qy         16 IleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeu 35
              ||||||   |||   |||:::::|||||||||||||||||
Db      29020 ATTATCCCCATTTTCCAGATGAAAAAACTGAGGCACAGAGCTAAAAGGCAGGATTCAAAC 28961

Qy         36 ProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyr 50
              ||||||   ||||||||||||:::||||||      ||||||:::
Db      28960 CCAGGTAGTCTGGCCCCAGAGTCAGTGCTG------TTTTGCTTC 28922

RESULT 10
AAX13160/c
ID   AAX13160 standard; DNA; 24601 BP.
XX
AC   AAX13160;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:223.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US008985.
XX
PR   06-MAY-1997;   97US-0044031P.
PR   16-MAY-1997;   97US-0046655P.
PR   14-NOV-1997;   97US-0066009P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA,  Dillon PJ,  Barash SC;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides -
PT   used to develop products for the detection of Enterococcus and for use in
PT   vaccines for prevention or attenuation of Enterococcus infection.
XX
PS   Claim 1; Page 1125-1137; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence of
CC   Enterococcus faecalis in samples. They can also be used for diagnosing
```

```
CC   Enterococcal infection in an animal and monitoring progression of
CC   disease, and for identifying agents which can be used to modulate the
CC   growth or pathogenicity of Enterococcus faecalis, or another related
CC   organism, in vivo or in vitro. In particular the polypeptides encoded by
CC   the Enterococcus faecalis nucleotide sequences can be used in vaccines to
CC   prevent or attenuate an Enterococcal infection
XX
SQ   Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
Pred. No.:              472          Length:        24601
Score:                  82.00        Matches:       42
Percent Similarity:     47.10%       Conservative:  23
Best Local Similarity:  30.43%       Mismatches:    51
Query Match:            11.82%       Indels:        22
DB:                     2            Gaps:          8

US-09-974-546C-86 (1-135) x AAX13160 (1-24601)

Qy          7 AsnGlnLysTyrGluAsp--------------------MetHisAsnIleIle------ 17
              |||||||||   :::                       |||::::::|||
Db       2571 AACCAAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTTCACGACTTAATAAAACAC 2512

Qy         18 ---------------HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
                            |||   |||:::   |||   ::::::
Db       2511 GCTTAAAATAAGGCACATTCACTAAAAGAAAGGCCAGTAAAATAATTAAGCCCACCCGTT 2452

Qy         33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
              |||   |||       |||:::   |||          |||||||||      |||   :::
Db       2451 CCATCATTAAAATAAATAATTCGACCACACCGGCACCTCCTTCCTATTAACTATACTATA 2392

Qy         53 PheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSer 72
                 ::::::|||||||||   :::   |||||||||         ||||||   :::|||
Db       2391 AATCAAAGAAAAGAAAAAAATGAAGCTTCTCAAAAA---AGAGGCTTCATT---TTTTCT 2338

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              |||         :::       ||||||   ||||||:::       |||
Db       2337 ATT---CATTAAAACGCAGGAATATTGGCGCCTTTAGAATTTTTCTTGATGTATTCAGCA 2281

Qy         93 SerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThr------ 110
              :::|||             :::   |||   :::   ::::::|||:::|||:::
Db       2280 ACTTCTTTTGTTGTCATCGCT---TCTTTTAAAACCTTGATTTTTTCGCTATCTTTTTCG 2224

Qy        111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
                 |||::::::|||   |||           |||   |||      ||||||
Db       2223 CCTTTGCGTGTCACA---ACTTGCAATGCGTAATGTTTACTTGCTTCTTTATCT 2173

RESULT 11
ABS98955/c
ID   ABS98955 standard; DNA; 24601 BP.
XX
AC   ABS98955;
XX
DT   18-DEC-2002  (first entry)
XX
DE   Enterococcus faecalis contig sequence #223.
XX
KW   Computer readable medium; Enterococcus faecalis; microbe; growth;
KW   pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
KW   therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
KW   biochip technology; antibacterial; modulator of nucleic acid expression;
KW   contig; ds.
XX
OS   Enterococcus faecalis.
XX
PN   US2002120116-A1.
XX
PD   29-AUG-2002.
XX
PF   04-MAY-1998;   98US-00070927.
XX
PR   04-MAY-1998;   98US-00070927.
XX
PA   (KUNS/) KUNSCH C A.
PA   (DILL/) DILLON P J.
PA   (BARA/) BARASH S.
XX
PI   Kunsch CA,  Dillon PJ,  Barash S;
XX
DR   WPI; 2002-750065/81.
XX
PT   Computer readable medium having recorded on it a Enterococcus faecalis
PT   nucleotide sequence useful for detecting diseases related to Enterococcus
PT   infections in animals.
XX
PS   Claim 1; Page; 119pp; English.
XX
CC   The present invention relates to a new computer readable medium with an
CC   Enterococcus faecalis nucleotide sequence. The invention is useful to
CC   diagnose the presence of E.faecalis in a sample or determining the
CC   presence of a specific microbe in a sample. The invention is also useful
CC   for modulating the growth or pathogenicity of E.faecalis, in a vaccine to
CC   confer resistance to Enterococcal infection, for commercial, therapeutic
CC   and industrial purposes, and for fermenting a particular sugar source or
CC   to produce a particular metabolite. The invention is useful for detecting
CC   diseases related to Enterococcus infections in animals, and for detecting
CC   E.faecalis using biochip technology. The present nucleic acid sequence
CC   represents an Enterococcus faecalis contig DNA sequence of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at http.sequdata.uspto.gov
XX
SQ   Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
Pred. No.:              472          Length:        24601
Score:                  82.00        Matches:       42
Percent Similarity:     47.10%       Conservative:  23
Best Local Similarity:  30.43%       Mismatches:    51
Query Match:            11.82%       Indels:        22
DB:                     6            Gaps:          8

US-09-974-546C-86 (1-135) x ABS98955 (1-24601)

Qy          7 AsnGlnLysTyrGluAsp--------------------MetHisAsnIleIle------ 17
              |||||||||   :::                       |||::::::|||
Db       2571 AACCAAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTTCACGACTTAATAAAACAC 2512

Qy         18 ---------------HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
                            |||   |||:::   |||   ::::::
Db       2511 GCTTAAAATAAGGCACATTCACTAAAAGAAAGGCCAGTAAAATAATTAAGCCCACCCGTT 2452

Qy         33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
              |||   |||       |||:::   |||          |||||||||      |||   :::
Db       2451 CCATCATTAAAATAAATAATTCGACCACACCGGCACCTCCTTCCTATTAACTATACTATA 2392

Qy         53 PheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSer 72
                 ::::::|||||||||   :::   |||||||||         ||||||   :::|||
Db       2391 AATCAAAGAAAAGAAAAAAATGAAGCTTCTCAAAAA---AGAGGCTTCATT---TTTTCT 2338

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              |||         :::       ||||||   ||||||:::       |||
Db       2337 ATT---CATTAAAACGCAGGAATATTGGCGCCTTTAGAATTTTTCTTGATGTATTCAGCA 2281

Qy         93 SerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThr------ 110
              :::|||             :::   |||   :::   ::::::|||:::|||:::
Db       2280 ACTTCTTTTGTTGTCATCGCT---TCTTTTAAAACCTTGATTTTTTCGCTATCTTTTTCG 2224

Qy        111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
                 |||::::::|||   |||           |||   |||      ||||||
Db       2223 CCTTTGCGTGTCACA---ACTTGCAATGCGTAATGTTTACTTGCTTCTTTATCT 2173

RESULT 12
ABD32622/c
ID   ABD32622 standard; DNA; 133462 BP.
XX
```

```
AC   ABD32622;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Mouse cancer-associated genomic DNA MD13-011.
XX
KW   Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Mus musculus.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   disclosure; seqid 140; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a mouse CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 133462 BP; 39091 A; 24305 C; 25093 G; 41557 T; 0 U; 3416 Other;

Alignment Scores:
Pred. No.:              6.32e+03        Length:         133462
```

```
Score:                  81.50           Matches:        37
Percent Similarity:     42.86%          Conservative:   17
Best Local Similarity:  29.37%          Mismatches:     33
Query Match:            11.74%          Indels:         39
DB:                     13              Gaps:           8

US-09-974-546C-86 (1-135) x ABD32622 (1-133462)

Qy         17 IleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuPro 36
              :::|||   |||           :::|||        :::||||||   :::
Db       9988 CTACATGCACTGCTAAATGGTGATGTCAGATTCGGTTTTACTAACTTCTCCAAA------ 9935

Qy         37 GlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLys 56
                           ||||||      |||   |||   |||:::         |||
Db       9934 --------------AATACTGTAAAGGGTCCTCTTTGCAGGAAGTTA---------AAA 9899

Qy         57 GluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSer 76
              :::|||   |||||||||   |||   :::           |||   :::      |||
Db       9898 AAAAAAAAAAAAAAGGTCGTTCAAAGTCAGTGCAATGGCTTCTTACTTCCTTAATTCAAGT 9839

Qy         77 HisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLys 96
              ::::::|||:::                                |||||||||   :::
Db       9838 AATTATGCTGTT----------------------------AGCGGTTCGTCT---CAA 9812

Qy         97 CysSerSerLeuSerSerGlu-----------------AlaIleLeuPheThrLeuThr 110
                 ||||||         :::                :::||||||   :::
Db       9811 TTTTCCTCTAAAGGAAAACAGTTTTAAACTATATAAATATTGCTCCTTTTTAATATCTTT 9752

Qy        111 LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysThrIle 130
              |||:::                                |||||   ||||||||||||
Db       9751 TTAAGG----------------------------AAGCTTTAT---TCAAAAACAATA 9725

Qy        131 HisPro---GlnIleIle 135
              ||||||   |||:::|||
Db       9724 CATCCCTCCCAAGTCATC 9707
```

```
RESULT 13
ABX44573
ID   ABX44573 standard; cDNA; 381 BP.
XX
AC   ABX44573;
XX
DT   21-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #9738.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-00960352.
XX
PR   12-JAN-1999;   99US-0115707P.
PR   11-JAN-2000; 2000US-00480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
```

```
XX
PS   Claim 2; SEQ ID NO 9738; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3' non- translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 381 BP; 104 A; 84 C; 82 G; 111 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.51          Length:          381
Score:                    81.00         Matches:         24
Percent Similarity:       42.55%        Conservative:    16
Best Local Similarity:    25.53%        Mismatches:      30
Query Match:              11.67%        Indels:          24
DB:                       8             Gaps:            4

US-09-974-546C-86 (1-135) x ABX44573 (1-381)

Qy         24 LysLeuArgHisArgLeuSerAsnPhePro------ArgLeuProGly------------ 37
              |||:::||||||:::::          |||      |||||||||
Db         96 AAGATAAGACACCAGATTCTCAAACTGCCAACTAGTCGTCTGCCCAGTAGCTGACGACGG 155

Qy         38 ------------------------IleLeuAlaProGluThrValLeuLeuProPheCys 49
                                     :::      |||         :::   |||
Db        156 AAAGGTTTGCCAACCCGTTCGCCCGTACACCTGCCCCGTTTCACGATGCAACCTGCTTCT 215

Qy         50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
                 |||   :::   :::::::::   |||                        ||||||
Db        216 GCAAAGTGGTACGATCGGAGGGACTATGTC-----------------------TTCATT 251

Qy         70 AspTyrSerIleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThr 89
              ::::::   :::|||   |||   |||      :::   ||||||::::::   ||||||
Db        252 GAATTTTGTGTCGAAGACAGTCAACATGTTAATGTAGATCCTTTAAAATCGAAACTTACA 311

Qy         90 MetLys------GlySerSerMetLysCysSerSerLeuSer 101
                  |||   |||      |||      |||:::
Db        312 TTCAGTTGTCTTGGAGGAAGTGATAATTGTAAACATTTAAAT 353

RESULT 14
ADN46845_11/c
Continuation (12 of 21) of ADN46845 from base 1100001 (Thermococcus kodakaraensis KOD1 g
WP Sequence split into 21 fragments  LOCUS ADN46845 Accession Adn46845
WP     Fragment Name      Begin     End
WP     ADN46845_00            1    110000
WP     ADN46845_01       100001    210000
WP     ADN46845_02       200001    310000
WP     ADN46845_03       300001    410000
WP     ADN46845_04       400001    510000
WP     ADN46845_05       500001    610000
WP     ADN46845_06       600001    710000
WP     ADN46845_07       700001    810000
WP     ADN46845_08       800001    910000
WP     ADN46845_09       900001   1010000
WP     ADN46845_10      1000001   1110000
WP     ADN46845_11      1100001   1210000
WP     ADN46845_12      1200001   1310000
WP     ADN46845_13      1300001   1410000
WP     ADN46845_14      1400001   1510000
WP     ADN46845_15      1500001   1610000
WP     ADN46845_16      1600001   1710000
WP     ADN46845_17      1700001   1810000
WP     ADN46845_18      1800001   1910000
WP     ADN46845_19      1900001   2010000
WP     ADN46845_20      2000001   2089378

Alignment Scores:
Pred. No.:                5.52e+03      Length:          110000
Score:                    81.00         Matches:         31
Percent Similarity:       44.36%        Conservative:    28
Best Local Similarity:    23.31%        Mismatches:      42
Query Match:              11.67%        Indels:          32
DB:                       12            Gaps:            7

US-09-974-546C-86 (1-135) x ADN46845_11 (1-110000)

Qy         13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer------ 30
              ||||||   :::         :::   ::::::||||||   |||   :::|||
Db      68597 ATGCACCAGCTT------GCAATCTATCTGAAAAAGCTCGGCCACGACGTTTCAATAGTG 68544

Qy         31 ------------------------------------------------AsnPhePro 33

Db      68543 ACTAATGATCTAAAAACGGGAAAGGAAAAAGAGCTTGAAGAGCTGGGGGTAGGATTGGTT 68484

Qy         34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
              ::::::||||||:::::::::|||      :::   :::       |||
Db      68483 AAAGTCCCGGGTGTCATCAGTCCG---GTGCTCGGGATAAACATAACCTAC--------- 68436

Qy         54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72
                           :::|||   :::::::::      ||||||   :::|||:::
Db      68435 -----------GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTTCTCGTGGATTTTGAT 68388

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              :::      :::|||||||||      |||||||||                     |||
Db      68387 GTTGTCCACGCCCACCACGCC---TTCACTCCACTCTCGCTGAAGGCGGTTAAGGCCGGG 68331

Qy         93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
              ::::::   |||   :::   |||::::::   :::|||   |||:::
Db      68330 AGAACTCTCGAAAAGGCAACACTCCTCACAACCCACAGCATATCCTTCTCCCACGAGTCT 68271

Qy        112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124
              |||   :::   |||||||||          ||||||
Db      68270 TCCCTATGGAAGGCCCTTGGGTTGACTTTTCCACTCCTC 68232

RESULT 15
ADN47591_09
Continuation (10 of 21) of ADN47591 from base 900001 (Thermococcus kodakaraensis KOD1 ger
WP Sequence split into 21 fragments  LOCUS ADN47591 Accession Adn47591
WP     Fragment Name      Begin     End
WP     ADN47591_00            1    110000
WP     ADN47591_01       100001    210000
WP     ADN47591_02       200001    310000
WP     ADN47591_03       300001    410000
WP     ADN47591_04       400001    510000
WP     ADN47591_05       500001    610000
WP     ADN47591_06       600001    710000
WP     ADN47591_07       700001    810000
WP     ADN47591_08       800001    910000
WP     ADN47591_09       900001   1010000
WP     ADN47591_10      1000001   1110000
WP     ADN47591_11      1100001   1210000
WP     ADN47591_12      1200001   1310000
WP     ADN47591_13      1300001   1410000
```

```
WP      ADN47591_14       1400001   1510000
WP      ADN47591_15       1500001   1610000
WP      ADN47591_16       1600001   1710000
WP      ADN47591_17       1700001   1810000
WP      ADN47591_18       1800001   1910000
WP      ADN47591_19       1900001   2010000
WP      ADN47591_20       2000001   2089378

Alignment Scores:
Pred. No.:              5.52e+03        Length:         110000
Score:                  81.00           Matches:        31
Percent Similarity:     44.36%          Conservative:   28
Best Local Similarity:  23.31%          Mismatches:     42
Query Match:            11.67%          Indels:         32
DB:                     12              Gaps:           7

US-09-974-546C-86 (1-135) x ADN47591_09 (1-110000)

Qy         13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer------ 30
              ||||||   :::         :::   ::::::||||||   |||   :::|||
Db      20781 ATGCACCAGCTT------GCAATCTATCTGAAAAAGCTCGGCCACGACGTTTCAATAGTG 20834

Qy         31 --------------------------------------------------AsnPhePro 33

Db      20835 ACTAATGATCTAAAAACGGGAAAGGAAAAAGAGCTTGAAGAGCTGGGGGTAGGATTGGTT 20894

Qy         34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
              ::::::||||||:::::::::|||      :::   :::         |||
Db      20895 AAAGTCCCGGGTGTCATCAGTCCG---GTGCTCGGGATAAACATAACCTAC--------- 20942

Qy         54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72
                        :::|||   :::::::::      ||||||   :::|||:::
Db      20943 ------------GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTTCTCGTGGATTTTGAT 20990

Qy         73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
              :::      :::|||||||||      |||||||||                  |||
Db      20991 GTTGTCCACGCCCACCACGCC---TTCACTCCACTCTCGCTGAAGGCGGTTAAGGCCGGG 21047

Qy         93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                ::::::   |||   :::   |||::::::   :::|||   |||:::
Db      21048 AGAACTCTCGAAAAGGCAACACTCCTCACAACCCACAGCATATCCTTCTCCCACGAGTCT 21107

Qy        112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124
                 |||   :::   |||||||||         ||||||
Db      21108 TCCCTATGGAAGGCCCTTGGGTTGACTTTTCCACTCCTC 21146


Search completed: September  9, 2005, 13:25:08
Job time : 388 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:12:05 ; Search time 104.5 Seconds
(without alignments)
2113.850 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174503_21006/app_query.fasta_1.654
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_77_@runat_07092005_174503_21006 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	694	100.0	2087	3	US-09-097-199-83	Sequence 83, Appl
	2	694	100.0	2505	3	US-09-097-199-85	Sequence 85, Appl
	3	694	100.0	2506	4	US-09-949-016-3311	Sequence 3311, Ap
	4	694	100.0	6507	4	US-09-949-016-15053	Sequence 15053, A
c	5	689	99.3	601	4	US-09-949-016-119365	Sequence 119365,
c	6	689	99.3	601	4	US-09-949-016-119366	Sequence 119366,
c	7	480	69.2	601	4	US-09-949-016-119364	Sequence 119364,
	8	83.5	12.0	2472	4	US-09-248-796A-148	Sequence 148, App
c	9	80.5	11.6	1093	3	US-09-186-276B-53	Sequence 53, Appl
c	10	80.5	11.6	1093	3	US-08-842-445-53	Sequence 53, Appl
c	11	80.5	11.6	1093	3	US-09-186-188B-53	Sequence 53, Appl
c	12	80.5	11.6	1093	4	US-09-265-585C-53	Sequence 53, Appl
c	13	80	11.5	148783	4	US-09-949-016-15729	Sequence 15729, A
c	14	79.5	11.5	262	4	US-09-573-080A-99	Sequence 99, Appl
	15	79.5	11.5	6196	3	US-08-675-566-5	Sequence 5, Appli
	16	79.5	11.5	6243	3	US-08-675-566-14	Sequence 14, Appl
	17	79.5	11.5	6503	3	US-08-675-566-6	Sequence 6, Appli
	18	79.5	11.5	7379	3	US-08-675-566-13	Sequence 13, Appl
	19	79.5	11.5	8618	3	US-08-675-566-21	Sequence 21, Appl
	20	79.5	11.5	8792	3	US-08-675-566-25	Sequence 25, Appl
	21	78	11.2	761	4	US-09-270-767-1826	Sequence 1826, Ap
	22	78	11.2	761	4	US-09-270-767-17108	Sequence 17108, A
	23	78	11.2	96878	4	US-09-949-016-12551	Sequence 12551, A
c	24	78	11.2	462589	4	US-09-949-016-12900	Sequence 12900, A
c	25	78	11.2	476044	4	US-09-949-016-12412	Sequence 12412, A
	26	77	11.1	1461	1	US-08-587-670A-1	Sequence 1, Appli
	27	77	11.1	1461	3	US-09-061-674-1	Sequence 1, Appli
	28	77	11.1	22372	4	US-09-949-016-17459	Sequence 17459, A
	29	76	11.0	367	3	US-09-328-111-446	Sequence 446, App
c	30	76	11.0	601	4	US-09-949-016-202994	Sequence 202994,
	31	76	11.0	42672	4	US-09-949-016-17253	Sequence 17253, A
	32	76	11.0	42672	4	US-09-949-016-17254	Sequence 17254, A
	33	76	11.0	66213	4	US-09-949-016-11803	Sequence 11803, A
	34	76	11.0	66213	4	US-09-949-016-16739	Sequence 16739, A
c	35	75.5	10.9	19008	4	US-09-949-016-12923	Sequence 12923, A
c	36	75.5	10.9	160759	4	US-09-949-016-16514	Sequence 16514, A
	37	75	10.8	64291	4	US-09-949-016-16278	Sequence 16278, A
	38	75	10.8	117410	4	US-09-949-016-12262	Sequence 12262, A
	39	74.5	10.7	3915	4	US-09-023-655-1104	Sequence 1104, Ap
c	40	74.5	10.7	5522	4	US-09-949-016-1259	Sequence 1259, Ap
c	41	74.5	10.7	5523	4	US-09-949-016-1014	Sequence 1014, Ap
	42	74.5	10.7	27120	4	US-09-949-016-16210	Sequence 16210, A
	43	74.5	10.7	35707	4	US-09-949-016-17120	Sequence 17120, A
c	44	74	10.7	601	4	US-09-949-016-157153	Sequence 157153,
c	45	74	10.7	3143	4	US-09-949-016-856	Sequence 856, App

ALIGNMENTS

RESULT 1
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:

```
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  83:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2087 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..503
US-09-097-199-83

Alignment Scores:
Pred. No.:              1.65e-85       Length:         2087
Score:                  694.00         Matches:        135
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-974-546C-86 (1-135) x US-09-097-199-83 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
US-09-097-199-85
; Sequence 85, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  85:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  99..503
US-09-097-199-85

Alignment Scores:
Pred. No.:              2.21e-85       Length:         2505
Score:                  694.00         Matches:        135
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-974-546C-86 (1-135) x US-09-097-199-85 (1-2505)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 3
US-09-949-016-3311
; Sequence 3311, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 14:46:00 ; Search time 408.5 Seconds
(without alignments)
2170.324 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174504_21077/app_query.fasta_1.654
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09974546_@CGN_1_1_456_@runat_07092005_174504_21077
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	2087	10	US-09-974-546-83	Sequence 83, Appl
2	694	100.0	2505	10	US-09-974-546-85	Sequence 85, Appl
3	90	13.0	1113	17	US-10-369-493-44085	Sequence 44085, A
4	90	13.0	1146	17	US-10-369-493-25021	Sequence 25021, A
5	87	12.5	1258	18	US-10-425-114-318	Sequence 318, App
6	87	12.5	1624	20	US-10-425-115-49918	Sequence 49918, A
c 7	83	12.0	146547	15	US-10-017-128-1	Sequence 1, Appli
c 8	82	11.8	24601	9	US-09-070-927A-223	Sequence 223, App
c 9	81.5	11.7	133462	19	US-10-367-094-114	Sequence 114, App
10	81	11.7	381	9	US-09-960-352-9738	Sequence 9738, Ap
11	81	11.7	819	19	US-10-437-963-23499	Sequence 23499, A
c 12	81	11.7	1277	20	US-10-425-115-137953	Sequence 137953,
13	81	11.7	4437	20	US-10-425-115-178334	Sequence 178334,
c 14	80.5	11.6	455	20	US-10-425-115-63753	Sequence 63753, A
c 15	80.5	11.6	600	22	US-10-972-079-4637	Sequence 4637, Ap
c 16	80.5	11.6	600	22	US-10-972-079-4638	Sequence 4638, Ap
c 17	80.5	11.6	1026	18	US-10-425-114-13022	Sequence 13022, A
c 18	80.5	11.6	1093	9	US-09-186-276B-53	Sequence 53, Appl
c 19	80.5	11.6	1093	9	US-09-186-188B-53	Sequence 53, Appl
c 20	80.5	11.6	1093	14	US-10-253-007-53	Sequence 53, Appl
21	80.5	11.6	2343	20	US-10-425-115-127057	Sequence 127057,
22	80	11.5	681	19	US-10-437-963-98199	Sequence 98199, A
23	80	11.5	1355	19	US-10-437-963-37754	Sequence 37754, A
24	80	11.5	1693	20	US-10-723-860-7296	Sequence 7296, Ap
c 25	79.5	11.5	262	10	US-09-854-867-99	Sequence 99, Appl
c 26	79.5	11.5	262	21	US-10-786-970A-99	Sequence 99, Appl
27	79.5	11.5	567	18	US-10-424-599-74400	Sequence 74400, A
28	79.5	11.5	585	20	US-10-357-930-54224	Sequence 54224, A
29	79	11.4	590	13	US-10-027-632-269701	Sequence 269701,
30	79	11.4	590	17	US-10-027-632-269701	Sequence 269701,
31	79	11.4	831	13	US-10-027-632-133030	Sequence 133030,
32	79	11.4	831	17	US-10-027-632-133030	Sequence 133030,
c 33	79	11.4	4068	19	US-10-437-963-35166	Sequence 35166, A
c 34	79	11.4	4107	20	US-10-425-115-139445	Sequence 139445,
35	79	11.4	4335	20	US-10-425-115-139432	Sequence 139432,
c 36	78.5	11.3	386	9	US-09-960-352-10443	Sequence 10443, A
37	78.5	11.3	600	22	US-10-972-079-88625	Sequence 88625, A
c 38	78.5	11.3	1787	13	US-10-027-632-84669	Sequence 84669, A
c 39	78.5	11.3	1787	13	US-10-027-632-84670	Sequence 84670, A
c 40	78.5	11.3	1787	17	US-10-027-632-84669	Sequence 84669, A
c 41	78.5	11.3	1787	17	US-10-027-632-84670	Sequence 84670, A
c 42	78	11.2	650	13	US-10-027-632-223621	Sequence 223621,
c 43	78	11.2	650	17	US-10-027-632-223621	Sequence 223621,
44	78	11.2	1207	18	US-10-424-599-5755	Sequence 5755, Ap
c 45	78	11.2	9869	20	US-10-425-115-139430	Sequence 139430,

ALIGNMENTS

RESULT 1
US-09-974-546-83
; Sequence 83, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```
;               STATE: Texas
;               COUNTRY: USA
;               ZIP: 77210
;          COMPUTER READABLE FORM:
;               MEDIUM TYPE: Floppy disk
;               COMPUTER: IBM PC compatible
;               OPERATING SYSTEM: PC-DOS/MS-DOS
;               SOFTWARE: PatentIn Release #1.0, Version #1.30
;          CURRENT APPLICATION DATA:
;               APPLICATION NUMBER: US/09/974,546
;               FILING DATE: 10-Oct-2001
;               CLASSIFICATION: Unknown
;          PRIOR APPLICATION DATA:
;               APPLICATION NUMBER: 09/097,199
;               FILING DATE: 1998-06-12
;          ATTORNEY/AGENT INFORMATION:
;               NAME: Nakashima, Richard A.
;               REGISTRATION NUMBER: P-42,023
;               REFERENCE/DOCKET NUMBER: UROC:018
;          TELECOMMUNICATION INFORMATION:
;               TELEPHONE: (512) 418-3000
;               TELEFAX: (512) 474-7577
;     INFORMATION FOR SEQ ID NO: 83:
;          SEQUENCE CHARACTERISTICS:
;               LENGTH: 2087 base pairs
;               TYPE: nucleic acid
;               STRANDEDNESS: single
;               TOPOLOGY: linear
;          FEATURE:
;               NAME/KEY:  CDS
;               LOCATION:  99..503
;          SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-974-546-83

Alignment Scores:
Pred. No.:                5.82e-85        Length:         2087
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-86 (1-135) x US-09-974-546-83 (1-2087)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 AGAAGTCAAAAGGCAACAGAGTTCATTGATTATTCCATAGAACAGTCACACCATGCAATT 338

Qy         81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 CTCACACCCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCCTCATTA 398

Qy        101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCTTCAGAAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCTAGGTCTGGAA 458

Qy        121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
              |||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGCTGTCTTCTCTACTTATCCAAAACTATACATCCACAGATCATA 503

RESULT 2
```

```
US-09-974-546-85
; Sequence 85, Application US/09974546
; Publication No. US20030050470A1
;  GENERAL INFORMATION:
;       APPLICANT: An, Gang
;                  O'Hara, S. Mark
;                  Ralph, David
;                  Veltri, Robert
;       TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                           PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Arnold, White & Durkee
;            STREET: P.O. Box 4433
;            CITY: Houston
;            STATE: Texas
;            COUNTRY: USA
;            ZIP: 77210
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/974,546
;            FILING DATE: 10-Oct-2001
;            CLASSIFICATION: Unknown
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/097,199
;            FILING DATE: 1998-06-12
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Nakashima, Richard A.
;            REGISTRATION NUMBER: P-42,023
;            REFERENCE/DOCKET NUMBER: UROC:018
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (512) 418-3000
;            TELEFAX: (512) 474-7577
;  INFORMATION FOR SEQ ID NO: 85:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2505 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  99..503
;       SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-974-546-85

Alignment Scores:
Pred. No.:                7.74e-85        Length:         2505
Score:                    694.00          Matches:        135
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-974-546C-86 (1-135) x US-09-974-546-85 (1-2505)

Qy          1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGAGGGCCTTCTTAAGGAACCAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158

Qy         21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218

Qy         41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAAGAAAAAGTAAAA 278

Qy         61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:09:35 ; Search time 1867.5 Seconds
(without alignments)
2751.634 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09974546/runat_07092005_174502_20989/app_query.fasta_1.654
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546_@CGN_1_1_3556_@runat_07092005_174502_20989 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	97	14.0	690	5	BU242169	BU242169 603779592
	2	95.5	13.8	588	5	BW569427	BW569427 BW569427
	3	94	13.5	606	6	CB286029	CB286029 CMD29_B11
	4	90.5	13.0	459	5	BW557636	BW557636 BW557636
	5	90.5	13.0	543	5	BW582739	BW582739 BW582739
	6	90	13.0	570	5	BW591794	BW591794 BW591794
c	7	89.5	12.9	1151	5	BU413225	BU413225 603155489
c	8	89	12.8	811	5	BU436282	BU436282 603209819
	9	88.5	12.8	743	7	CV511381	CV511381 kc40h08.y
	10	88.5	12.8	789	8	BH559290	BH559290 BOHOW80TR
	11	88	12.7	776	9	AG377362	AG377362 Mus muscu
c	12	87.5	12.6	803	5	BX756993	BX756993 BX756993
c	13	87	12.5	788	9	CR262699	CR262699 Reverse s
	14	86.5	12.5	741	3	AY439721	AY439721 Armigeres
c	15	86	12.4	529	5	BQ592283	BQ592283 E012698-0
	16	86	12.4	769	8	BZ521554	BZ521554 BOKAE11TF
c	17	86	12.4	992	9	CL254017	CL254017 ZMMBBb060
c	18	85.5	12.3	788	8	CC091268	CC091268 CSU-K33r.
c	19	85.5	12.3	877	6	CA472302	CA472302 AGENCOURT
c	20	85	12.2	579	9	CE464906	CE464906 tigr-gss-
c	21	85	12.2	707	9	CC743728	CC743728 ZMMBBb011
	22	85	12.2	906	9	CG300818	CG300818 OG0AB25TV
c	23	85	12.2	914	9	CG300807	CG300807 OG0AB25TH
	24	84.5	12.2	285	4	BM111157	BM111157 EST558693
	25	84.5	12.2	562	8	AZ954412	AZ954412 2M0220005
c	26	84.5	12.2	652	2	BF644675	BF644675 NF018F01E
c	27	84.5	12.2	688	5	BP027905	BP027905 BP027905
c	28	84.5	12.2	741	4	BI309709	BI309709 EST531119
c	29	84.5	12.2	770	4	BM780176	BM780176 EST590764
	30	84	12.1	493	5	BX836326	BX836326 BX836326
	31	84	12.1	624	8	BZ745987	BZ745987 OGFBK05TM
c	32	84	12.1	728	6	CD821188	CD821188 BN25.041B
	33	84	12.1	896	9	CG008627	CG008627 ZUACO85TH
	34	84	12.1	916	7	CK151729	CK151729 FGAS03441
c	35	84	12.1	969	2	BE739923	BE739923 601556175
c	36	84	12.1	1399	2	AW730867	AW730867 GA__Ea002
	37	83.5	12.0	721	8	B69065	B69065 CIT-HSP-205
	38	83.5	12.0	732	5	BU264381	BU264381 603505585
	39	83.5	12.0	781	7	CN135272	CN135272 OX1_31_E0
	40	83.5	12.0	952	1	AL665968	AL665968 AL665968
	41	83.5	12.0	1335	4	BM467342	BM467342 AGENCOURT
	42	83	12.0	395	6	CB069753	CB069753 is28d02.x
	43	83	12.0	572	2	BE015616	BE015616 EST842 Ma
	44	83	12.0	612	2	BE015598	BE015598 EST824 Ma
	45	83	12.0	634	9	AG110445	AG110445 Pan trogl

ALIGNMENTS

RESULT 1
BU242169/c
LOCUS BU242169 690 bp mRNA linear EST 26-NOV-2002
DEFINITION 603779592F1 CSEQCHN34 Gallus gallus cDNA clone ChEST725h23 5', mRNA sequence.
ACCESSION BU242169
VERSION BU242169.1 GI:25488447
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 690)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES Location/Qualifiers
source 1. .690
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 15I"

```
                     /db_xref="taxon:9031"
                     /clone="ChEST725h23"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN34"
                     /note="Organ: liver; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

Alignment Scores:
Pred. No.:                1.02          Length:         690
Score:                    97.00         Matches:        33
Percent Similarity:       45.08%        Conservative:   22
Best Local Similarity:    27.05%        Mismatches:     59
Query Match:              13.98%        Indels:         8
DB:                       5             Gaps:           3

US-09-974-546C-86 (1-135) x BU242169 (1-690)

Qy         15 AsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArg 34
              |||:::      ||||||:::        :::          :::||||||
Db        623 AACTTAAATAGCATTTTACGTTTCCCTACAGTTACTAAAGAAATATCAAAT------GAT 570

Qy         35 LeuProGlyIleLeuAlaPro---------GluThrValLeuLeuProPheCysTyrLys 51
              :::||||||:::|||   |||                    ||||||
Db        569 GTCCCAGGACTCCTGTTGCCCACAGAGGAGAACAGGTATTTGCTCAGCACTATCACAGTG 510

Qy         52 ValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyr 71
                ||||||   |||   :::          :::   ||||||   |||
Db        509 AAATTTAGGTGAAAATAAAGAACAGTATACACTGATAAAGCTGCAGAG---------AAT 459

Qy         72 SerIleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLys 91
              |||:::   :::   |||   :::      :::      :::      |||   ::::::
Db        458 TCAGTTATAAAAGGCCATTTCTCATTTCAGAGCTTTGCTAGGATGCCTCTTCTTATAAGA 399

Qy         92 GlySerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
                        :::   :::         :::|||   :::||||||
Db        398 AAAAAAGGAAAAAGAAAAGCACCACCACAAACTGAGATGCTGCTATTTCTCACTAAATGC 339

Qy        112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysThrIleHis 131
              :::   |||   |||             ||||||||||||:::         |||||||||
Db        338 AAGGCCACCACTACATCCTCTTGTTTCTGTTGCCTTCTGTTTACAGGAGTAACCATTCAT 279

Qy        132 ProGln 133
              |||:::
Db        278 CCAAAG 273

RESULT 2
BW569427
LOCUS       BW569427               588 bp    mRNA    linear   EST 31-AUG-2004
DEFINITION  BW569427 Yutaka Satou unpublished cDNA library (cstb) Ciona
            savignyi cDNA clone cstb014a08 3', mRNA sequence.
ACCESSION   BW569427
VERSION     BW569427.1  GI:51749861
KEYWORDS    EST.
SOURCE      Ciona savignyi
  ORGANISM  Ciona savignyi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 588)
  AUTHORS   Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona savignyi
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Yutaka Satou
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4095
            Fax: 81-75-705-1113
            Email: yutaka@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .588
                     /organism="Ciona savignyi"
                     /mol_type="mRNA"
                     /db_xref="taxon:51511"
                     /clone="cstb014a08"
                     /dev_stage="tailbud stage"
                     /clone_lib="Yutaka Satou unpublished cDNA library (cstb)"
ORIGIN

Alignment Scores:
Pred. No.:                1.23          Length:         588
Score:                    95.50         Matches:        37
Percent Similarity:       44.27%        Conservative:   21
Best Local Similarity:    28.24%        Mismatches:     48
Query Match:              13.76%        Indels:         25
DB:                       5             Gaps:           6

US-09-974-546C-86 (1-135) x BW569427 (1-588)

Qy         24 LysLeuArgHisArgLeuSerAsn-----------ProArgLeuProGlyIleLeu 39
              |||      |||:::|||   |||            |||   :::|||      :::
Db         82 AAATATTCACATAAACTTGGCAACAAAAACACAAGATTTTTACAACTTGCTCTAGTTTGA 141

Qy         40 AlaProGluThrValLeuLeuProPheCysTyrLysValPhe-----------ArgLys 55
                 |||::::::   |||            |||::::::
Db        142 CTAGAAGAATCCATTTTTTTAAAGGACGAAAGCAAAATTTACATAATTTGCAATAATCCG 201

Qy         56 LysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGln 75
                |||   :::::::|||   ::::::   ::::::|||::::::||||||   ||||||
Db        202 TAGGAATTGCTTAGAAGATTTAAAAGAAGAAGCGATTTCGTAGAATATAGCACGGAACAA 261

Qy         76 Ser------HisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySer 93
              |||      ||||||||||||         :::|||:::   |||   :::   ||||||
Db        262 AGCAATTTGCACCACGCTATATTTGCAGGGATTCAGTCATTTCTTCCACTGAGTGGCTCA 321

Qy         94 SerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeu 113
                    |||          |||            |||                 ::::::
Db        322 AGAGAGCACTGTTGCTTGTGCAGCTGCTGATACTGCCTC-----------CAGCGGATT 369

Qy        114 ThrGlnThrLeuGlyLeuGluCys--------------------CysLeuLeu------ 124
                    :::||||||   |||                        |||   |||
Db        370 GGGACTGGGGTGGGTCTGTTGTGCATTCGTCAAGAATTGTGAGGGTGCTCGTTGGTCCAC 429

Qy        125 ------TyrLeuSerLysThrIleHisProGln 133
                    ||||||            ||||||
Db        430 TACTGCTACCTTGTGGATGCTTTTCCTCCACAA 462

RESULT 3
CB286029
LOCUS       CB286029                606 bp    mRNA    linear   EST 27-FEB-2003
DEFINITION  CMD29_B11_90 UMNMPM3 Sus scrofa cDNA clone PPSUBLIB_15B11 5', mRNA
            sequence.
ACCESSION   CB286029
VERSION     CB286029.1  GI:28576983
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Dvorak,C.M.T., Hyland,K.A., Zhang,Y., Fahrenkrug,S.C. and
            Murtaugh,M.P.
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:23:48 ; Search time 58 Seconds
(without alignments)
900.218 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	3	AAY59296	Aay59296 Prostate
2	694	100.0	135	3	AAY59295	Aay59295 Prostate
3	694	100.0	135	4	AAU02174	Aau02174 Biomarker
4	694	100.0	135	4	AAU02175	Aau02175 Biomarker
5	103	14.8	21	3	AAY59294	Aay59294 Cancer bi
6	103	14.8	21	4	AAU02173	Aau02173 Biomarker
7	90	13.0	371	8	ADS41968	Ads41968 Bacterial
8	90	13.0	381	8	ADN18681	Adn18681 Bacterial
9	82.5	11.9	645	5	ABB90843	Abb90843 Herbicida
10	81	11.7	384	8	ADN48099	Adn48099 Thermococ
11	77.5	11.2	1061	7	ADF50279	Adf50279 Human PFM
12	77.5	11.2	1061	8	ADO36741	Ado36741 Human PR-
13	77.5	11.2	1144	4	ABB58124	Abb58124 Drosophil
14	77	11.1	1026	8	ADS08184	Ads08184 Staphyloc
15	76.5	11.0	322	6	ABM69043	Abm69043 Photorhab
16	75	10.8	633	6	ABU42279	Abu42279 Protein e
17	73	10.5	195	4	ABG18410	Abg18410 Novel hum
18	72	10.4	523	4	AAB46398	Aab46398 H. pylori
19	72	10.4	636	8	ADM94292	Adm94292 Soybean E
20	71.5	10.3	450	7	ADF41696	Adf41696 Bacillus
21	71.5	10.3	634	8	ADM94288	Adm94288 Corn Ethy
22	71.5	10.3	676	4	ABG15498	Abg15498 Novel hum
23	71	10.2	127	4	AAU42328	Aau42328 Propionib
24	71	10.2	127	6	ABM38847	Abm38847 Propionib
25	71	10.2	424	4	AAM39676	Aam39676 Human pol
26	71	10.2	424	4	AAG81358	Aag81358 Human AFP
27	71	10.2	424	8	ADH09599	Adh09599 Human hos
28	71	10.2	424	8	ADH09598	Adh09598 Human hos
29	71	10.2	424	8	ADH09600	Adh09600 Human hos
30	71	10.2	429	4	AAM41462	Aam41462 Human pol
31	71	10.2	519	4	AAB92584	Aab92584 Human pro
32	71	10.2	1922	8	ADL97802	Adl97802 Human dcr
33	71	10.2	1924	4	AAE09768	Aae09768 Human dic
34	71	10.2	1924	5	AAO15987	Aao15987 Human hel
35	71	10.2	1924	6	ABO07204	Abo07204 Human p53
36	71	10.2	1924	7	ABU64479	Abu64479 Stem loop
37	71	10.2	1924	7	ADN60159	Adn60159 Human hel
38	70.5	10.2	438	7	ABO64633	Abo64633 Klebsiell
39	70.5	10.2	634	8	ADM94300	Adm94300 Corn Ethy
40	70	10.1	206	7	ADJ69892	Adj69892 Human hea
41	70	10.1	982	8	ADS29846	Ads29846 Bacterial
42	69.5	10.0	1294	2	AAW30601	Aaw30601 Human typ
43	69.5	10.0	1353	8	ADQ89904	Adq89904 Antagonis
44	69.5	10.0	1398	3	AAB18292	Aab18292 Plasmodiu
45	69.5	10.0	3105	5	ABB80604	Abb80604 Human sbg

ALIGNMENTS

RESULT 1
AAY59296
ID AAY59296 standard; peptide; 135 AA.
XX
AC AAY59296;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AAZ87584.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 184-186; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used

```
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 135 AA;

  Query Match             100.0%;  Score 694;  DB 3;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-75;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 2
AAY59295
ID   AAY59295 standard; peptide; 135 AA.
XX
AC   AAY59295;
XX
DT   19-APR-2000  (first entry)
XX
DE   Prostate disease marker UC Band #28 amino acid sequence.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
DR   N-PSDB; AAZ87583.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Example 5; Page 182-183; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 135 AA;

  Query Match             100.0%;  Score 694;  DB 3;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-75;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 3
AAU02174
ID   AAU02174 standard; protein; 135 AA.
XX
AC   AAU02174;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker protein encoded by UC band 28 #2, used in diagnosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   N-PSDB; AAS04000.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Disclosure; Col 121; 78pp; English.
XX
CC   The sequence represents the amino acid sequence of biomarker protein
CC   encoded by UC band 28 #2. Proteins encoded by the nucleic acid markers
CC   can be used to produce antibodies for the detection of prostate, breast
CC   or bladder cancer. Biomarker nucleic acid sequences can be used as
CC   hybridisation probes and primers that specifically hybridise to prostate
CC   cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
CC   cancer markers. The nucleic acids can be used as targets for therapeutic
CC   intervention in these diseases, in the identification and isolation of
CC   full-length gene sequences, including regulatory elements for gene
CC   expression, from genomic human DNA libraries, as hybridisation probes for
CC   screening genomic human DNA libraries. The kits comprising the nucleic
CC   acid sequences are useful for detecting bladder, breast or prostate
```

```
CC   cancer cells in a biological sample
XX
SQ   Sequence 135 AA;

  Query Match             100.0%;  Score 694;  DB 4;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-75;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 4
AAU02175
ID   AAU02175 standard; protein; 135 AA.
XX
AC   AAU02175;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker protein encoded by UC band 28 #3, used in diagnosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
DR   N-PSDB; AAS04001.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Disclosure; Col 125; 78pp; English.
XX
CC   The sequence represents the amino acid sequence of biomarker protein
CC   encoded by UC band 28 #3. Proteins encoded by the nucleic acid markers
CC   can be used to produce antibodies for the detection of prostate, breast
CC   or bladder cancer. Biomarker nucleic acid sequences can be used as
CC   hybridisation probes and primers that specifically hybridise to prostate
CC   cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
CC   cancer markers. The nucleic acids can be used as targets for therapeutic
CC   intervention in these diseases, in the identification and isolation of
CC   full-length gene sequences, including regulatory elements for gene
CC   expression, from genomic human DNA libraries, as hybridisation probes for
CC   screening genomic human DNA libraries. The kits comprising the nucleic
CC   acid sequences are useful for detecting bladder, breast or prostate
CC   cancer cells in a biological sample
XX
SQ   Sequence 135 AA;

  Query Match             100.0%;  Score 694;  DB 4;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-75;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 5
AAY59294
ID   AAY59294 standard; peptide; 21 AA.
XX
AC   AAY59294;
XX
DT   19-APR-2000  (first entry)
XX
DE   Cancer biomarker UC28 antigenic peptide.
XX
KW   Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW   benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS   Homo sapiens.
XX
PN   WO9964631-A1.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US013151.
XX
PR   12-JUN-1998;   98US-00097199.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri RW;
XX
DR   WPI; 2000-116557/10.
XX
PT   Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT   breast and bladder cancer.
XX
PS   Example 5; Page 114; 191pp; English.
XX
CC   The invention provides nucleic acid markers of prostate, breast and
CC   bladder cancer. The markers are indicators of malignant transformation of
CC   prostate, breast and bladder tissues and are diagnostic of the potential
CC   for metastatic spread of malignant prostate tumours. The nucleic acid can
CC   also be used as targets for therapeutic intervention in prostate cancer,
CC   benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC   markers may be used to design specific probes and primers, for the rapid
CC   analysis of prostate, bladder or breast biopsy samples. The probes and
CC   primers may also be used for in situ hybridization or in situ PCR
CC   detection and diagnosis. They may also be used to identify and isolate
CC   full length gene sequences form various DNA libraries. Antibodies against
CC   the polypeptide products of the markers can be used to treat prostate
CC   cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC   to detect antibodies. The proteins and antibodies can be used in
CC   immunodetection methods for detecting or quantifying the cancers, and for
CC   clinical diagnosis of these cancers. The antibodies may also be used for
CC   radioimaging to quantify and localize the encoded proteins
XX
SQ   Sequence 21 AA;
```

```
  Query Match             14.8%;  Score 103;  DB 3;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-05;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         54 RKKEKVKRSQKATEFIDYSIE 74
              |||||||||||||||||||||
Db          1 RKKEKVKRSQKATEFIDYSIE 21


RESULT 6
AAU02173
ID   AAU02173 standard; peptide; 21 AA.
XX
AC   AAU02173;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Biomarker UC band 28, antigenic peptide used in diagnosis of cancer.
XX
KW   Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW   benign prostatic hyperplasia; BPH; therapeutic; human; antigenic.
XX
OS   Homo sapiens.
XX
PN   US6218529-B1.
XX
PD   17-APR-2001.
XX
PF   12-JUN-1998;   98US-00097199.
XX
PR   31-JUL-1995;   95US-0001655P.
PR   11-JAN-1996;   96US-0013611P.
PR   31-JUL-1996;   96US-00692787.
XX
PA   (UROC-) UROCOR INC.
XX
PI   An G,  O'hara SM,  Ralph D,  Veltri R;
XX
DR   WPI; 2001-289849/30.
XX
PT   New nucleic acids as biomarkers and targets useful for detecting,
PT   diagnosing, prognosing, and in developing treatments for prostate, breast
PT   and bladder cancer.
XX
PS   Example 5; Col 73; 78pp; English.
XX
CC   The sequence represents the amino acid sequence of biomarker, UC band 28,
CC   antigenic peptide used to produce antibodies for the detection of
CC   prostate, breast or bladder cancer. Biomarker nucleic acid sequences can
CC   be used as hybridisation probes and primers that specifically hybridise
CC   to prostate cancer, benign prostatic hyperplasia (BPH), bladder cancer or
CC   breast cancer markers. The nucleic acids can be used as targets for
CC   therapeutic intervention in these diseases, in the identification and
CC   isolation of full-length gene sequences, including regulatory elements
CC   for gene expression, from genomic human DNA libraries, as hybridisation
CC   probes for screening genomic human DNA libraries. The kits comprising the
CC   nucleic acid sequences are useful for detecting bladder, breast or
CC   prostate cancer cells in a biological sample
XX
SQ   Sequence 21 AA;

  Query Match             14.8%;  Score 103;  DB 4;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-05;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         54 RKKEKVKRSQKATEFIDYSIE 74
              |||||||||||||||||||||
Db          1 RKKEKVKRSQKATEFIDYSIE 21


RESULT 7
ADS41968
ID   ADS41968 standard; protein; 371 AA.
XX
AC   ADS41968;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #20398.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 20398; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 371 AA;

  Query Match             13.0%;  Score 90;  DB 8;  Length 371;
  Best Local Similarity   26.8%;  Pred. No. 0.1;
  Matches   34;  Conservative   23;  Mismatches   38;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRL----SNFP--------------RLPGILAPETVLLPFCYKVF 53
```

```
              |||:   :::|:  | :    :| |               ::|||::|        :
Db         21 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 70

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                :| |::  ||: |: |  ||||  |||       | :| | : |:: :| |
Db         71 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 129

Qy        112 QLTQTLG 118
              :|  |||
Db        130 KLWDTLG 136


RESULT 8
ADN18681
ID   ADN18681 standard; protein; 381 AA.
XX
AC   ADN18681;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #1334.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 1334; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 381 AA;

  Query Match             13.0%;  Score 90;  DB 8;  Length 381;
  Best Local Similarity   26.8%;  Pred. No. 0.11;
  Matches   34;  Conservative   23;  Mismatches   38;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRL----SNFP--------------RLPGILAPETVLLPFCYKVF 53
              |||:   :::|:  | :    :| |               ::|||::|        :
Db         24 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 73

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                :| |::  ||: |: |  ||||  |||       | :| | : |:: :| |
Db         74 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132

Qy        112 QLTQTLG 118
              :|  |||
Db        133 KLWDTLG 139


RESULT 9
ABB90843
ID   ABB90843 standard; protein; 645 AA.
XX
AC   ABB90843;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 54.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP009892.
XX
PR   28-AUG-2001; 2001WO-EP009892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms.
XX
PS   Claim 5; SEQ ID NO 54; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins (ABB90790-ABB94016)
CC   for herbicidally active compounds, comprising aligning and comparing
CC   nucleic acid or amino acid sequences from plant with nucleic acid or
CC   amino acid sequences from non-plant organisms using suitable search
CC   parameters, where plant sequences having an E-value greater by a factor
CC   of 3 than the E-value of most similar non-plant sequences are selected.
CC   The polypeptides or nucleic acids encoding them are useful for
CC   identifying modulators. The identified modulators are useful as
CC   herbicides
XX
SQ   Sequence 645 AA;
```

```
  Query Match             11.9%;  Score 82.5;  DB 5;  Length 645;
  Best Local Similarity   23.5%;  Pred. No. 1.7;
  Matches   31;  Conservative   26;  Mismatches   46;  Indels   29;  Gaps    5;

Qy         25 LRHRLSNF---PRLPGILAPETV-----------------LLPFCYKVFRKKEKVKRSQ 63
              : | |: |    |  | ::   ||                 |||   |   |  :   |:
Db         99 MTHLLAGFTYGPHWPWVMTAVTVFKMLTGIVSFLTALSLVTLLPLLLKA--KVREFMLSK 156

Qy         64 KATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKC-SSLSSEAILFTLTLQLTQTLGLECC 122
              |  |     :::    |:   :| | ::  : |  :||    ||:|  ::|::||||: |
Db        157 KTRE-----LDREVGIIMKQTETSLHVRMLTTKIRTSLDRHTILYTTLVELSKTLGLKNC 211

Qy        123 LLYLSKTIHPQI 134
               :::   |  ::
Db        212 AVWIPNEIKTEM 223


RESULT 10
ADN48099
ID   ADN48099 standard; protein; 384 AA.
XX
AC   ADN48099;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Thermococcus kodakaraensis KOD1 protein sequence SeqID1977.
XX
KW   gene disruption; gene targeting; marker gene; transformation;
KW   homologous recombination; hyperthermostable archaebacterium; KOD1;
KW   gene structure; gene function; enzyme activity; medicine;
KW   forensic science; food; drug inspection; molecular biology; immunology.
XX
OS   Thermococcus kodakaraensis.
XX
PN   WO2004022736-A1.
XX
PD   18-MAR-2004.
XX
PF   29-AUG-2003; 2003WO-IB003597.
XX
PR   30-AUG-2002; 2002JP-00319011.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Imanaka T,  Atomi H;
XX
DR   WPI; 2004-257583/24.
XX
PT   Method for disrupting targeted gene in genome of organism particularly
PT   thermostable bacterium and with genome chips for analysis, applicable in
PT   studying gene structure and functions.
XX
PS   Claim 9; SEQ ID NO 1977; 598pp; Japanese.
XX
CC   This invention relates to a novel method for targeting disruption of an
CC   arbitrary gene in a genome of an organism which comprises providing the
CC   whole sequential data of the genome of such organism, selecting at least
CC   1 arbitrary region in the sequence, providing a vector that contains a
CC   sequence homologous with the selected region and a marker gene,
CC   transformation, and homologous recombination. The genome is preferably
CC   the genome of a hyperthermostable archaebacterium, particularly
CC   Thermococcus kodakaraensis KOD1. The method is for targeting the
CC   disruption of a gene in the genome of an organism, which is applicable in
CC   studying gene structure and functions as well as enzyme activities of
CC   encoded proteins and useful in medicine, forensic science, food or drug
CC   inspection, molecular biology and immunology. With this method, the
CC   disruption of a gene at an arbitrary position in a genome can be achieved
CC   efficiently and reliably. The present sequence is that of a protein
CC   encoded by the genome of Thermococcus kodakaraensis which was derived
CC   using the method of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 384 AA;

  Query Match             11.7%;  Score 81;  DB 8;  Length 384;
  Best Local Similarity   23.3%;  Pred. No. 1.3;
  Matches   31;  Conservative   28;  Mismatches   42;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRLS------------------NFPRLPGILAPETVLLPFCYKVF 53
              || :   : ::|| | :|                  ::||:::|  : :   |
Db         21 MHQL--AIYLKKLGHDVSIVTNDLKTGKEKELEELGVGLVKVPGVISP-VLGINITY--- 74

Qy         54 RKKEKVKRSQKATEF-IDYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                    :| :::  || :|: :  :||| |||       | :: | : |:: :| |:
Db         75 ----GLKSNRELGEFLVDFDVVHAHHA-FTPLSLKAVKAGRTLEKATLLTTHSISFSHES 129

Qy        112 QLTQTLGLECCLL 124
               | : |||   ||
Db        130 SLWKALGLTFPLL 142


RESULT 11
ADF50279
ID   ADF50279 standard; protein; 1061 AA.
XX
AC   ADF50279;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human PFM7 protein, a PR/SET family member.
XX
KW   human; PFM7; SET; Su(var)3-9, Enhancer-of-Zeste and Trithorax; cancer;
KW   PFM/SET; tumour suppressor; hyperproliferative disorder; scleroderma;
KW   arthritis; alcoholic liver cirrhosis; hypertropic scarring;
KW   atherosclerosis; gene therapy; Duchenne's muscular dystrophy;
KW   insulin-dependent diabetes mellitus; Huntington's; Parkinson's;
KW   Alzheimer's disease; paralysis; cerebellar atrophy; cytostatic;
KW   neuroprotective; nootropic; antiarthritic; vulnerary;
KW   antiatherosclerotic; antidiabetic; hepatotropic; PFM; PR family member.
XX
OS   Homo sapiens.
XX
PN   US2003049623-A1.
XX
PD   13-MAR-2003.
XX
PF   18-JUL-2001; 2001US-00910478.
XX
PR   18-JUL-2001; 2001US-00910478.
XX
PA   (HUAN/) HUANG S.
XX
PI   Huang S;
XX
DR   WPI; 2003-567062/53.
DR   N-PSDB; ADF50278.
XX
PT   New PR/SET-domain containing nucleic acids (which encodes PFM/SET) and
PT   polypeptides, useful for preventing or treating cancers, scleroderma,
PT   arthritis, keloids, atherosclerosis, Huntington's disease or Alzheimer's
PT   disease.
XX
PS   Claim 21; SEQ ID NO 4; 83pp; English.
XX
CC   This invention relates to novel PFM (PR family member)/SET (Su(var)3-9,
CC   Enhancer-of-Zeste and Trithorax) nucleotides and encoded polypeptides
CC   thereof that are implicated in proliferative disorders such as cancer.
CC   Specifically, it refers to a functional fragment of the PFM/SET tumour
CC   suppressor gene that encodes a PR, SET, PRAZ or PKZL domain. The present
CC   invention describes the identification and characterisation of additional
CC   PR/SET-domain family members that can be used as regulators of cell
CC   proliferation and furthermore to treat, prevent or diagnose
CC   hyperproliferative disorders including scleroderma, arthritis, alcoholic
```

CC liver cirrhosis, hypertropic scarring and atherosclerosis. Through gene
CC therapy, these polynucleotides can be used to enhance proliferation of
CC normal cells without rendering the cells cancerous and as such they are
CC particularly useful for treating Duchenne's muscular dystrophy, insulin-
CC dependent diabetes mellitus, Huntington's, Parkinson's, Alzheimer's
CC disease, paralysis, or cerebellar atrophy. Accordingly, these
CC compositions can be described as cytostatic, neuroprotective, nootropic,
CC antiarthritic, vulnerary, antiatherosclerotic, antidiabetic or
CC hepatotropic. This polypeptide sequence is the human PFM7 protein
CC (encoded by a PFM/SET gene localised to chromosome 11q25) of the
CC invention.
XX
SQ Sequence 1061 AA;

Query Match 11.2%; Score 77.5; DB 7; Length 1061;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 17; Mismatches 42; Indels 31; Gaps 6;

```
Qy         18 HILQ----------IRKLR--------HRLSNFPRLPGILAPETVLLPFCYKVFRKKEKV 59
              |||:          |||||          ||   :| | :|   |  | | | :  | |:
Db        704 HILKNHPGAELPPSIRKLRPAGPGEPDPMLSTHTQLTGTIATPPVCCPHCSKQYSSKTKM 763

Qy         60 KR--SQKATEFIDYSIEQSHHAILTPLQTH-LTMKGSSMKCSSLSSEAILFTLTL----- 111
               :   :|  ||   |     : | ||| |  ::   : :   | : | :: |  |
Db        764 VQHIRKKHPEFAQLS-----NTIHTPLTTAVISATPAVLTTDSATGETVVTTDLLTQAMT 818

Qy        112 QLTQTL 117
              :|:|||
Db        819 ELSQTL 824
```

RESULT 12
ADO36741
ID ADO36741 standard; protein; 1061 AA.
XX
AC ADO36741;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human PR-domain containing protein PFM-7.
XX
KW cytostatic; histone methyltransferase; gene therapy;
KW PFM/SET protein binding motif; cell growth modulator;
KW histone methyltransferase activity; proliferative disorder; cancer;
KW PR-domain; PFM-7.
XX
OS Homo sapiens.
XX
PN US2004014192-A1.
XX
PD 22-JAN-2004.
XX
PF 18-JUL-2002; 2002US-00200012.
XX
PR 18-JUL-2002; 2002US-00200012.
XX
PA (HUAN/) HUANG S.
XX
PI Huang S;
XX
DR WPI; 2004-121568/12.
XX
PT New isolated nucleic acid molecule comprising a sequence encoding a
PT PFM/SET polypeptide, useful for diagnosing, prognosing, preventing and
PT treating proliferative disorders, e.g. cancer.
XX
PS Claim 1; SEQ ID NO 4; 38pp; English.
XX
CC The invention describes an isolated nucleic acid molecule comprising a
CC sequence encoding a PFM/SET protein binding motifs polypeptides
CC comprising a 689, 1061, 367, 717, 571, 951, 1257, 720, or 770 amino acid
CC sequence (P1-P9), given in the specification. Also described are: a
CC vector comprising the isolated nucleic acid molecule; a host cell
CC comprising the vector; an oligonucleotide comprising at least 17
CC contiguous nucleotides of the nucleic acid molecule; a primer pair for
CC detecting PFM/SET nucleic acid molecule, comprising two isolated
CC oligonucleotides; detecting PFM/SET nucleic acid molecule in a sample;
CC modulating cell growth by introducing the vector into a host cell, and
CC expressing the encoded PFM/SET polypeptide in an amount effective to
CC modulate growth of the cell; isolating PFM/SET polypeptide by growing the
CC host cell under conditions appropriate for the expression of the
CC polypeptide; the isolated PFM/SET polypeptide, or its functional fragment
CC ; an isolated immunogenic PFM/SET peptide, comprising at least 8
CC contiguous amino acids of P1-P9; an antibody or its antigen-binding
CC fragment that specifically binds to PFM/SET polypeptide; detecting
CC PFM/SET polypeptide in a sample; and screening for a compound that
CC modulates PFM/SET polypeptide histone methyltransferase activity by
CC contacting the polypeptide with one or more candidate compounds, and
CC determining histone methyltransferase activity of the contacted
CC polypeptide. The nucleic acid molecules, polypeptides and methods are
CC useful for diagnosing, prognosing, preventing and treating proliferative
CC disorders, e.g. cancer. This is the amino acid sequence of PR-domain
CC containing polypeptide PFM-7.
XX
SQ Sequence 1061 AA;

Query Match 11.2%; Score 77.5; DB 8; Length 1061;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 17; Mismatches 42; Indels 31; Gaps 6;

```
Qy         18 HILQ----------IRKLR--------HRLSNFPRLPGILAPETVLLPFCYKVFRKKEKV 59
              |||:          |||||          ||   :| | :|   |  | | | :  | |:
Db        704 HILKNHPGAELPPSIRKLRPAGPGEPDPMLSTHTQLTGTIATPPVCCPHCSKQYSSKTKM 763

Qy         60 KR--SQKATEFIDYSIEQSHHAILTPLQTH-LTMKGSSMKCSSLSSEAILFTLTL----- 111
               :   :|  ||   |     : | ||| |  ::   : :   | : | :: |  |
Db        764 VQHIRKKHPEFAQLS-----NTIHTPLTTAVISATPAVLTTDSATGETVVTTDLLTQAMT 818

Qy        112 QLTQTL 117
              :|:|||
Db        819 ELSQTL 824
```

RESULT 13
ABB58124
ID ABB58124 standard; protein; 1144 AA.
XX
AC ABB58124;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 1164.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02227.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell

```
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 1164; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1144 AA;

  Query Match             11.2%;  Score 77.5;  DB 4;  Length 1144;
  Best Local Similarity   30.0%;  Pred. No. 15;
  Matches   30;  Conservative   12;  Mismatches   33;  Indels   25;  Gaps    5;

Qy         15 NIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKK------EKVKRSQKATEF 68
              | :| || ||||| ||          : : ||  :| |  |      || |   :  :|
Db        718 NQVHELQ-RKLRHVSSNL---------KAIDLPTNFKFFFLKTDRHGQEKAK--SQIQKF 765

Qy         69 IDYSIEQSH-------HAILTPLQTHLTMKGSSMKCSSLS 101
              ::: :|  |       :  |:|   ||     | | |  |
Db        766 LNFILEDDHLNGSEAIYTFLSPSSDHLKQSLPSPKKSKFS 805


RESULT 14
ADS08184
ID   ADS08184 standard; protein; 1026 AA.
XX
AC   ADS08184;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Staphylococcus epidermis polypeptide seqid 7479.
XX
KW   antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW   recombinant expression vector; infection; computer readable medium;
KW   computer based system.
XX
OS   Staphylococcus epidermidis.
XX
PN   US2004147734-A1.
XX
PD   29-JUL-2004.
XX
PF   01-DEC-2003; 2003US-00724972.
XX
PR   08-NOV-1997;   97US-0064964P.
PR   13-AUG-1998;   98US-00134001.
PR   29-NOV-1999;   99US-00450969.
XX
PA   (DOUC/) DOUCETTE-STAMM L.
PA   (BUSH/) BUSH D.
XX
PI   Doucette-Stamm L,  Bush D;
XX
DR   WPI; 2004-580138/56.
DR   N-PSDB; ADS04412.
XX
PT   New isolated polypeptide and encoding nucleic acid derived from
PT   Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT   treating an S. epidermidis bacterial infection.
XX
PS   Claim 17; SEQ ID NO 7479; 741pp; English.
XX
CC   The invention describes an isolated nucleic acid comprising a nucleotide
CC   sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC   1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC   of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC   given in the specification. Also described are: a recombinant expression
CC   vector; a cell comprising a recombinant expression vector of (1);
CC   producing an S. epidermidis polypeptide; an isolated nucleic acid
CC   comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC   vaccine composition for prevention or treatment of an S. epidermidis
CC   infection, comprising a nucleic acid cited above and a carrier; treating
CC   a subject for S. epidermidis infection; a recombinant or substantially
CC   pure preparation of an S. epidermidis polypeptide or its fragment; a
CC   vaccine composition for prevention or treatment of an S. epidermidis
CC   infection; detecting the presence of a Staphylococcus nucleic acid in a
CC   sample; a computer readable medium having recorded in it the nucleotide
CC   sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC   system for identifying fragments of the Staphylococcus genome of
CC   commercial importance; a computer based system for identifying fragments
CC   of the Staphylococcus plasmids of commercial importance; identifying
CC   commercially important nucleic acid fragments of the Staphylococcus
CC   genome and/or plasmids; and identifying an expression modulating fragment
CC   of the Staphylococcus genome and/or plasmids. The methods and
CC   compositions of the present invention are useful for the diagnosis,
CC   prevention and/or treatment of an Staphylococcal epidermidis bacterial
CC   infection. This is the amino acid sequence of a S. epidermis protein of
CC   the invention.
XX
SQ   Sequence 1026 AA;

  Query Match             11.1%;  Score 77;  DB 8;  Length 1026;
  Best Local Similarity   20.6%;  Pred. No. 15;
  Matches   35;  Conservative   33;  Mismatches   54;  Indels   48;  Gaps    8;

Qy          7 NQKYED-MHNIIHILQIR---------KLRHRLSNFPRLPGILAPETVL----LPFCYKV 52
              ||  || || :| |:::          |::  ||       :: | : |    ||  | :
Db        782 NQALEDEMHTVIKIVELSRQARKNADLKIKQPLSKM-----VIKPNSQLNLSFLPNYYSI 836

Qy         53 FRKKEKVKR---SQKATEFIDYSIEQSHHAI----------------LTPLQTHLTMKG 92
                : :  :|    :    ::| | :: :  ::               |:       ::
Db        837 IKDELNIKNIELTDNINDYITYELKLNFSSVGPKLGNKTKNIQTLIDSLSEYDKKSLIES 896

Qy         93 SSMKCSSLSSEA-------ILFTLTLQLTQTLGLECCLLYLSKTIHPQII 135
              :: |  ||||:|       |: ||      |      |:: | | : |::|
Db        897 NNFK--SLSSDAELTKDDFIIKTLPKDSYQLSEDNDCVILLDKNLSPELI 944


RESULT 15
ABM69043
ID   ABM69043 standard; protein; 322 AA.
XX
AC   ABM69043;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Photorhabdus luminescens protein sequence #2140.
XX
KW   Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW   detection; food; gene expression; plant; animal; microorganism; toxin;
KW   antibiotic; biopesticide; virulence factor; disease model; plague;
KW   whooping cough.
XX
OS   Photorhabdus luminescens.
XX
PN   WO200294867-A2.
XX
PD   28-NOV-2002.
XX
PF   07-FEB-2002; 2002WO-IB003040.
XX
PR   07-FEB-2001; 2001FR-00001659.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Duchaud E,  Taourit S,  Glaser P,  Frangeul L,  Kunst F,  Danchin A;
PI   Buchrieser C;
```

XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 2140; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 322 AA;

Query Match 11.0%; Score 76.5; DB 6; Length 322;
Best Local Similarity 26.6%; Pred. No. 3.5;
Matches 29; Conservative 19; Mismatches 50; Indels 11; Gaps 3;

```
Qy         13 MHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVKRSQKATEFIDYS 72
              :|| || :||   :  ||    ||||      ::   :|          ||| | | :
Db        221 IHNFIHAVQIPAEKFGLSRTVNLPGISVTVQEMIDALAEV--------AGQKAVELIRFE 272

Qy         73 IEQSHHAILTPLQTHLTM-KGSSM--KCSSLSSEAILFTLTLQLTQTLG 118
               ::: : |:     |  : :| |:     :  |: |   :|  |:|| |
Db        273 PDENINRIVASWPGHFDISRGLSLGFHADNTFSDTIRAFITNNLSQTGG 321
```

Search completed: September 9, 2005, 13:00:52
Job time : 59.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:25:33 ; Search time 17.5 Seconds
(without alignments)
575.864 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	3	US-09-097-199-84	Sequence 84, Appl
2	694	100.0	135	3	US-09-097-199-86	Sequence 86, Appl
3	694	100.0	157	4	US-09-949-016-9182	Sequence 9182, Ap
4	103	14.8	21	3	US-09-097-199-56	Sequence 56, Appl
5	72	10.4	636	4	US-09-614-912-92	Sequence 92, Appl
6	71.5	10.3	634	4	US-09-614-912-88	Sequence 88, Appl
7	70.5	10.2	438	4	US-09-489-039A-11150	Sequence 11150, A
8	70.5	10.2	634	4	US-09-614-912-100	Sequence 100, App
9	69.5	10.0	1294	3	US-09-473-717-2	Sequence 2, Appli
10	69.5	10.0	1294	4	US-09-949-016-6023	Sequence 6023, Ap
11	69.5	10.0	1353	3	US-09-398-193-99	Sequence 99, Appl
12	68.5	9.9	2470	4	US-08-265-967C-2	Sequence 2, Appli
13	68.5	9.9	2470	4	US-08-305-790B-3	Sequence 3, Appli
14	68	9.8	496	4	US-10-146-704-3	Sequence 3, Appli
15	68	9.8	639	4	US-09-949-016-6812	Sequence 6812, Ap
16	68	9.8	652	4	US-09-949-016-7323	Sequence 7323, Ap
17	68	9.8	993	4	US-09-949-016-10335	Sequence 10335, A
18	67.5	9.7	156	4	US-09-543-681A-7434	Sequence 7434, Ap
19	67	9.7	399	4	US-09-543-681A-4933	Sequence 4933, Ap
20	67	9.7	517	3	US-09-457-040B-14	Sequence 14, Appl
21	67	9.7	773	2	US-08-484-101B-42	Sequence 42, Appl
22	67	9.7	773	2	US-08-484-101B-44	Sequence 44, Appl
23	67	9.7	773	3	US-08-714-524D-42	Sequence 42, Appl
24	67	9.7	773	3	US-08-714-524D-44	Sequence 44, Appl
25	66.5	9.6	580	4	US-09-107-532A-5734	Sequence 5734, Ap
26	66.5	9.6	1784	3	US-09-040-738-2	Sequence 2, Appli
27	66.5	9.6	1784	3	US-08-652-426A-2	Sequence 2, Appli
28	66.5	9.6	1813	4	US-09-949-016-8283	Sequence 8283, Ap
29	66.5	9.6	1813	4	US-09-949-016-8284	Sequence 8284, Ap
30	66.5	9.6	1813	4	US-09-949-016-8285	Sequence 8285, Ap
31	66	9.5	267	4	US-09-270-767-46386	Sequence 46386, A
32	66	9.5	348	4	US-09-198-452A-922	Sequence 922, App
33	66	9.5	348	4	US-09-438-185A-856	Sequence 856, App
34	66	9.5	611	4	US-09-949-016-7139	Sequence 7139, Ap
35	65.5	9.4	698	4	US-09-538-092-151	Sequence 151, App
36	65.5	9.4	3218	1	US-08-764-100-27	Sequence 27, Appl
37	65	9.4	75	4	US-09-621-976-6155	Sequence 6155, Ap
38	65	9.4	258	4	US-09-270-767-33067	Sequence 33067, A
39	65	9.4	258	4	US-09-270-767-48284	Sequence 48284, A
40	65	9.4	288	1	US-08-375-709-9	Sequence 9, Appli
41	65	9.4	288	1	US-08-752-929-9	Sequence 9, Appli
42	65	9.4	288	3	US-09-090-793-6	Sequence 6, Appli
43	65	9.4	288	4	US-09-231-899-6	Sequence 6, Appli
44	64.5	9.3	236	4	US-09-107-532A-6873	Sequence 6873, Ap
45	64.5	9.3	315	3	US-09-134-001C-5446	Sequence 5446, Ap

ALIGNMENTS

```
RESULT 1
US-09-097-199-84
; Sequence 84, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
;   APPLICANT: An, Gang
;   APPLICANT: O'Hara, S. Mark
;   APPLICANT: Ralph, David
;   APPLICANT: Veltri, Robert
;   TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;   TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;   NUMBER OF SEQUENCES: 87
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Arnold, White & Durkee
;     STREET: P.O. Box 4433
;     CITY: Houston
;     STATE: Texas
;     COUNTRY: USA
;     ZIP: 77210
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/097,199.
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/692,787
;     FILING DATE: 31-JUL-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Nakashima, Richard A.
;     REGISTRATION NUMBER: P-42,023
;     REFERENCE/DOCKET NUMBER: UROC:018
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (512) 418-3000
;     TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 84:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 135 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-097-199-84

  Query Match 100.0%; Score 694; DB 3; Length 135;
  Best Local Similarity 100.0%; Pred. No. 4.7e-77;
  Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 2
US-09-097-199-86
; Sequence 86, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/692,787
;      FILING DATE:  31-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nakashima, Richard A.
;      REGISTRATION NUMBER:  P-42,023
;      REFERENCE/DOCKET NUMBER:  UROC:018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  86:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-097-199-86

  Query Match             100.0%;  Score 694;  DB 3;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-77;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 3
US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9182
;   LENGTH: 157
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-9182

  Query Match             100.0%;  Score 694;  DB 4;  Length 157;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-77;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 82

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 142

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        143 CCLLYLSKTIHPQII 157


RESULT 4
US-09-097-199-56
; Sequence 56, Application US/09097199
; Patent No. 6218529
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gang
;    APPLICANT:  O'Hara, S. Mark
;    APPLICANT:  Ralph, David
;    APPLICANT:  Veltri, Robert
;    TITLE OF INVENTION:  BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;    TITLE OF INVENTION:  PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;    NUMBER OF SEQUENCES:  87
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/097,199
;      FILING DATE:
;      CLASSIFICATION:
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:59:09 ; Search time 52.5 Seconds
(without alignments)
1014.252 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 1777461

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	10	US-09-974-546-84	Sequence 84, Appl
2	694	100.0	135	10	US-09-974-546-86	Sequence 86, Appl
3	103	14.8	21	10	US-09-974-546-56	Sequence 56, Appl
4	90	13.0	371	15	US-10-369-493-20398	Sequence 20398, A
5	90	13.0	381	15	US-10-369-493-1334	Sequence 1334, Ap
6	81	11.7	1478	16	US-10-425-115-362997	Sequence 362997,
7	79	11.4	264	15	US-10-424-599-174106	Sequence 174106,
8	79	11.4	1368	16	US-10-425-115-324108	Sequence 324108,
9	79	11.4	1444	16	US-10-425-115-324095	Sequence 324095,
10	78	11.2	289	15	US-10-424-599-148597	Sequence 148597,
11	78	11.2	3229	16	US-10-425-115-324093	Sequence 324093,
12	77.5	11.2	1061	15	US-10-200-012-4	Sequence 4, Appli
13	77	11.1	1026	18	US-10-724-972A-7479	Sequence 7479, Ap
14	77	11.1	1766	16	US-10-425-115-324098	Sequence 324098,
15	75	10.8	633	15	US-10-282-122A-70203	Sequence 70203, A
16	75	10.8	1259	16	US-10-425-115-324122	Sequence 324122,
17	75	10.8	3749	16	US-10-425-115-324125	Sequence 324125,
18	74.5	10.7	388	15	US-10-424-599-169767	Sequence 169767,
19	74.5	10.7	404	18	US-10-954-778-97	Sequence 97, Appl
20	74	10.7	1353	16	US-10-425-115-262744	Sequence 262744,
21	74	10.7	1605	16	US-10-425-115-362998	Sequence 362998,
22	74	10.7	1676	16	US-10-425-115-238569	Sequence 238569,
23	74	10.7	1708	16	US-10-425-115-324103	Sequence 324103,
24	74	10.7	1752	16	US-10-425-115-238562	Sequence 238562,
25	74	10.7	1754	16	US-10-425-115-324104	Sequence 324104,
26	73	10.5	1143	16	US-10-425-115-238564	Sequence 238564,
27	73	10.5	1517	16	US-10-425-115-324099	Sequence 324099,
28	72.5	10.4	3365	16	US-10-739-930-6537	Sequence 6537, Ap
29	72	10.4	329	15	US-10-424-599-150082	Sequence 150082,
30	71.5	10.3	407	15	US-10-425-114-69882	Sequence 69882, A
31	71.5	10.3	426	16	US-10-425-115-327356	Sequence 327356,
32	71.5	10.3	676	15	US-10-424-599-267862	Sequence 267862,
33	71	10.2	334	15	US-10-425-114-63742	Sequence 63742, A
34	71	10.2	1214	16	US-10-425-115-243074	Sequence 243074,
35	71	10.2	1595	16	US-10-425-115-238559	Sequence 238559,
36	71	10.2	1715	16	US-10-425-115-324123	Sequence 324123,
37	71	10.2	1924	9	US-09-866-557A-2	Sequence 2, Appli
38	71	10.2	1924	11	US-09-858-862-2	Sequence 2, Appli
39	71	10.2	1924	14	US-10-055-797-2	Sequence 2, Appli
40	71	10.2	1924	15	US-10-350-798-2	Sequence 2, Appli
41	70.5	10.2	269	15	US-10-424-599-264790	Sequence 264790,
42	70.5	10.2	634	17	US-10-876-086-27	Sequence 27, Appl
43	70	10.1	206	16	US-10-408-765A-1698	Sequence 1698, Ap
44	70	10.1	272	16	US-10-437-963-125982	Sequence 125982,
45	70	10.1	327	15	US-10-425-114-63743	Sequence 63743, A

ALIGNMENTS

RESULT 1
US-09-974-546-84
; Sequence 84, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199
; FILING DATE: 1998-06-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023

```
;             REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 84:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 135 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

  Query Match             100.0%;  Score 694;  DB 10;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-72;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 2
US-09-974-546-86
; Sequence 86, Application US/09974546
; Publication No. US20030050470A1
;   GENERAL INFORMATION:
;         APPLICANT: An, Gang
;                    O'Hara, S. Mark
;                    Ralph, David
;                    Veltri, Robert
;         TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                             PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;         NUMBER OF SEQUENCES: 87
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/974,546
;             FILING DATE: 10-Oct-2001
;             CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/097,199
;             FILING DATE: 1998-06-12
;         ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 86:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 135 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

  Query Match             100.0%;  Score 694;  DB 10;  Length 135;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-72;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135


RESULT 3
US-09-974-546-56
; Sequence 56, Application US/09974546
; Publication No. US20030050470A1
;   GENERAL INFORMATION:
;         APPLICANT: An, Gang
;                    O'Hara, S. Mark
;                    Ralph, David
;                    Veltri, Robert
;         TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
;                             PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
;         NUMBER OF SEQUENCES: 87
;         CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Arnold, White & Durkee
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210
;         COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/974,546
;             FILING DATE: 10-Oct-2001
;             CLASSIFICATION: Unknown
;         PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/097,199
;             FILING DATE: 1998-06-12
;         ATTORNEY/AGENT INFORMATION:
;             NAME: Nakashima, Richard A.
;             REGISTRATION NUMBER: P-42,023
;             REFERENCE/DOCKET NUMBER: UROC:018
;         TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;    INFORMATION FOR SEQ ID NO: 56:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 21 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-974-546-56

  Query Match             14.8%;  Score 103;  DB 10;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         54 RKKEKVKRSQKATEFIDYSIE 74
```

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:01:39 ; Search time 14.5 Seconds
(without alignments)
895.811 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90	13.0	381	2	F71196	probable hexosyltr
2	82.5	11.9	645	2	F86174	hypothetical prote
3	80	11.5	495	1	MNXRSA	nonstructural prot
4	78.5	11.3	3147	2	T21328	hypothetical prote
5	78	11.2	523	2	D64555	conserved hypothet
6	74.5	10.7	172	2	A58456	X protein - human
7	74.5	10.7	263	2	A82069	hypothetical prote
8	72.5	10.4	796	2	T43782	hypothetical prote
9	71.5	10.3	227	2	B84040	ABC transporter (A
10	71.5	10.3	451	2	D70045	two-component sens
11	70.5	10.2	1064	2	T40751	isoleucyl-trna syn
12	70	10.1	404	2	T06761	hypothetical prote
13	70	10.1	982	1	A44831	phosphoenolpyruvat
14	70	10.1	982	2	AE2413	phosphoenolpyruvat
15	69.5	10.0	1398	2	H71606	hypothetical prote
16	69	9.9	154	2	T27983	hypothetical prote
17	69	9.9	743	2	T00634	hypothetical prote
18	69	9.9	1850	2	AC1917	serine/threonine k
19	68.5	9.9	503	1	S11338	steroid 11beta-mon
20	68.5	9.9	2470	2	S57085	1-phosphatidylinos
21	68	9.8	379	2	C75006	hypothetical prote
22	68	9.8	414	1	H64203	histidine-tRNA lig
23	68	9.8	496	2	A31986	glucose transporte
24	68	9.8	639	2	T13151	adapter protein CM
25	68	9.8	5126	2	S40450	ryanodine receptor
26	67.5	9.7	619	2	B87682	hypothetical prote
27	67.5	9.7	833	2	T28385	ORF MSV224 probabl
28	67.5	9.7	1278	2	A71609	probable secreted
29	67.5	9.7	3225	2	D81702	adherence factor T
30	67	9.7	126	2	T03757	hypothetical prote
31	67	9.7	213	2	E96506	hypothetical prote
32	67	9.7	317	2	B82440	hypothetical prote
33	67	9.7	347	2	T01452	translation initia
34	67	9.7	517	2	A54099	protein kinase Dar
35	67	9.7	587	2	S37927	hypothetical prote
36	66.5	9.6	244	2	AC3341	precorrin-2 C20-me
37	66.5	9.6	447	2	T21716	hypothetical prote
38	66.5	9.6	635	2	T06537	ethylene receptor
39	66.5	9.6	859	2	H70327	DNA mismatch repai
40	66.5	9.6	908	2	T22376	hypothetical prote
41	66.5	9.6	1784	2	A49420	tuberous sclerosis
42	66	9.5	344	2	D86597	outer membrane pro
43	66	9.5	344	2	H72027	major outer membra
44	66	9.5	446	2	A42029	transcription fact
45	66	9.5	472	2	F97147	probable Fe-S oxid

ALIGNMENTS

RESULT 1
F71196
probable hexosyltransferase (EC 2.4.1.-) PH1844 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: F71196
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71196
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-381 <KAW>
A;Cross-references: UNIPROT:O59512; GB:AP000007; NID:g3236134; PIDN:BAA30965.1; PID:g3258
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1844
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 13.0%; Score 90; DB 2; Length 381;
Best Local Similarity 26.8%; Pred. No. 0.16;
Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;

```
Qy         13 MHNIIHILQIRKLRHRL----SNFP---------------RLPGILAPETVLLPFCYKVF 53
              |||:   :::|:  | :    :| |               ::|||::|        :
Db         24 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 73

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                 :| |::  ||: |: |  ||||  |||       | :| | : |:: :| |
Db         74 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132

Qy        112 QLTQTLG 118
                :|  |||
Db        133 KLWDTLG 139
```

RESULT 2
F86174
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: F86174
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-645 <STO>
A;Cross-references: UNIPROT:P93825; GB:AE005172; NID:g2341041; PIDN:AAB70445.1; GSPDB:GN
C;Genetics:
A;Map position: 1

Query Match 11.9%; Score 82.5; DB 2; Length 645;
Best Local Similarity 23.5%; Pred. No. 1.7;
Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;

```
Qy          25 LRHRLSNF---PRLPGILAPETV-----------------LLPFCYKVFRKKEKVKRSQ 63
               : | |: |   |  | ::   ||                  |||   |   |  :   |:
Db          99 MTHLLAGFTYGPHWPWVMTAVTVFKMLTGIVSFLTALSLVTLLPLLLKA--KVREFMLSK 156

Qy          64 KATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKC-SSLSSEAILFTLTLQLTQTLGLECC 122
               |  |     :::    |:   :| | ::  : |  :||    ||:|  ::|::||||: |
Db         157 KTRE-----LDREVGIIMKQTETSLHVRMLTTKIRTSLDRHTILYTTLVELSKTLGLKNC 211

Qy         123 LLYLSKTIHPQI 134
                :::   |  ::
Db         212 AVWIPNEIKTEM 223
```

RESULT 3
MNXRSA
nonstructural protein NCVP2 - simian rotavirus SA11
N;Alternate names: nonstructural protein NS53
C;Species: simian rotavirus SA11
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S08215
R;Mitchell, D.B.; Both, G.W.
Virology 174, 618-621, 1990
A;Title: Conservation of a potential metal binding motif despite extensive sequence dive
A;Reference number: S08215; MUID:90163231; PMID:2154894
A;Accession: S08215
A;Molecule type: genomic RNA
A;Residues: 1-495 <MIT>
A;Cross-references: UNIPROT:P15687; EMBL:X14914; NID:g61889; PIDN:CAA33039.1; PID:g61890
C;Genetics:
A;Map position: segment 5
C;Superfamily: bovine rotavirus nonstructural protein NCVP2
C;Keywords: nonstructural protein; zinc finger

Query Match 11.5%; Score 80; DB 1; Length 495;
Best Local Similarity 25.3%; Pred. No. 2.3;
Matches 24; Conservative 22; Mismatches 33; Indels 16; Gaps 4;

```
Qy           2 RAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVKR 61
               | || |:          |:|::| ::|  ::   :|  |:     ::  |   || || |:: :
Db          71 RCFLDNEP--------HLLKLRTVKHPITK-DKLQCIIDLYNIIFPINDKVIRKFERMIK 121

Qy          62 SQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMK 96
               :|      | ||  :| :|      :|:  :: |
Db         122 QRKCRN--QYKIEWYNHLLLP-----ITLNAAAFK 149
```

RESULT 4
T21328
hypothetical protein F25C8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21328
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19407
A;Accession: T21328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3147 <WIL>
A;Cross-references: UNIPROT:Q9XV66; EMBL:Z81512; PIDN:CAB04172.1; GSPDB:GN00023; CESP:F25
A;Experimental source: clone F25C8
C;Genetics:
A;Gene: CESP:F25C8.3
A;Map position: 5
A;Introns: 32/2; 57/2; 73/3; 113/3; 157/3; 366/1; 456/1; 509/1; 638/1; 773/2; 848/1; 902/
/3; 2510/2; 2737/3; 2827/1; 2902/3; 3093/3; 3121/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F25C8.3

Query Match 11.3%; Score 78.5; DB 2; Length 3147;
Best Local Similarity 28.1%; Pred. No. 26;
Matches 41; Conservative 17; Mismatches 67; Indels 21; Gaps 6;

```
Qy           2 RAFLRN-------QKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFR 54
               ||  ||         | : :   ||  |   :  | |      |   | : : :    |: :
Db        1571 RADKRNLLADNWQAKQQALRKSIHARQSTAVPRRESAMVGQPE-FASKAIKMMLMEKMQQ 1629

Qy          55 KKEKVKRSQKATE--FIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQ 112
               :||| |  :|  :      |:|| | :  |     |  |    |           | :| ||
Db        1630 EKEKEKEKEKEEKDALKKQSVEQDHSSTDTEEDAQLPEKNKPML-------TYLRSLVLQ 1682

Qy         113 LTQ---TLGLECCLLYLSKTIHPQII 135
               |     :  |:|||| ||   | |:|
Db        1683 LVHSPISSVLKCCLL-LSVEQHKQMI 1707
```

RESULT 5
D64555
conserved hypothetical integral membrane protein HP0284 - Helicobacter pylori (strain 26
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: D64555
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: D64555
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-523 <TOM>
A;Cross-references: UNIPROT:O25059; GB:AE000547; GB:AE000511; NID:g2313377; PIDN:AAD0735

Query Match 11.2%; Score 78; DB 2; Length 523;
Best Local Similarity 23.5%; Pred. No. 3.9;
Matches 24; Conservative 16; Mismatches 30; Indels 32; Gaps 3;

```
Qy           5 LRNQKYEDMHNIIHILQIRKLRHRLS----------------------NFPRLPGILA 40
               |:|:| :|      :|:|  | | |                          | | :| |
Db          71 LKNKKSKD------LLKISTLEHTLKALESQQKMFESYGVNPFKDLIERPNIPNIPNIAN 124

Qy          41 PETVL--LPFCYKVFRKKEKVKRSQKATEFIDYSIEQSHHAI 80
               |  ::  : |   :  | | :||:| | | :   ::| |  :
Db         125 PIAIIDGISFIKSMHLKHENLKRNQTALEEVLRLLDQKHQLL 166
```

RESULT 6
A58456
X protein - human hepatitis B virus
C;Species: human hepatitis B virus
C;Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 09-Jul-2004
C;Accession: A58456
R;Zeng, M.; Huang, B.R.; Cai, L.W.; Pan, G.Z.
Chinese Biochem. J. 12, 22-26, 1996
A;Title: The sequence analysis of the hepatitis B virus (HBV) X gene and its expression i
A;Reference number: JC1331
A;Accession: A58456
A;Molecule type: DNA
A;Residues: 1-172 <ZEN>

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:00:59 ; Search time 51.5 Seconds
(without alignments)
1342.343 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKYEDMHNIIHIL.....TLGLECCLLYLSKTIHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	694	100.0	135	2	Q9GZY1	Q9gzy1 homo sapien
2	90	13.0	381	2	O59512	O59512 pyrococcus
3	82.5	11.9	645	2	P93825	P93825 arabidopsis
4	81	11.7	451	2	Q7SZX8	Q7szx8 brachydanio
5	80	11.5	414	2	Q99FX6	Q99fx6 simian rota
6	80	11.5	479	2	Q99FX4	Q99fx4 simian rota
7	80	11.5	495	1	VN53_ROTS1	P15687 simian 11 r
8	80	11.5	496	2	Q99FX5	Q99fx5 simian rota
9	80	11.5	496	2	Q99FX7	Q99fx7 simian rota
10	78.5	11.3	3175	2	Q7JKT8	Q7jkt8 caenorhabdi
11	78	11.2	523	2	O25059	O25059 helicobacte
12	78	11.2	1646	2	Q86820	Q86820 potato aucu
13	77.5	11.2	358	2	Q8TZU8	Q8tzu8 pyrococcus
14	77.5	11.2	931	2	Q9NXI4	Q9nxi4 homo sapien
15	77.5	11.2	1023	2	Q863Z2	Q863z2 homo sapien
16	77.5	11.2	1061	1	PRDA_HUMAN	Q9nqv6 homo sapien
17	77.5	11.2	1061	2	Q9ULI9	Q9uli9 homo sapien
18	77.5	11.2	1117	2	Q8T4E8	Q8t4e8 drosophila
19	77.5	11.2	1117	2	Q9W3N0	Q9w3n0 drosophila
20	77	11.1	495	1	VN53_ROTSP	P35425 simian 11 r
21	77	11.1	1024	1	SYIP_STAAU	P41368 staphylococ
22	76.5	11.0	321	2	Q7N442	Q7n442 photorhabdu
23	76.5	11.0	398	2	Q6LRR8	Q6lrr8 photobacter
24	76.5	11.0	2259	1	YCF2_PHYPA	P61243 physcomitre
25	75	10.8	629	2	Q6GJ08	Q6gj08 staphylococ
26	74.5	10.7	160	2	Q6SIF8	Q6sif8 uncultured
27	74.5	10.7	163	2	Q6SCJ3	Q6scj3 uncultured
28	74.5	10.7	172	2	Q7M022	Q7m022 hepatitis b
29	74.5	10.7	263	2	Q9KP79	Q9kp79 vibrio chol
30	74.5	10.7	404	2	Q7PNX8	Q7pnx8 anopheles g
31	74	10.7	3184	2	Q9XV66	Q9xv66 caenorhabdi
32	73.5	10.6	615	2	Q9FHK4	Q9fhk4 arabidopsis
33	73.5	10.6	853	2	Q7R2I5	Q7r2i5 giardia lam
34	72.5	10.4	220	2	Q7P537	Q7p537 fusobacteri
35	72.5	10.4	220	2	Q8RF89	Q8rf89 fusobacteri
36	72.5	10.4	716	2	O21040	O21040 dictyosteli
37	72.5	10.4	796	2	Q9TGM3	Q9tgm3 dictyosteli
38	72.5	10.4	1283	2	O97034	O97034 ephydatia f
39	72.5	10.4	3306	2	Q9FT44	Q9ft44 arabidopsis
40	72	10.4	316	2	Q83DP2	Q83dp2 coxiella bu
41	72	10.4	400	2	Q6RX30	Q6rx30 arabidopsis
42	72	10.4	448	2	Q6ALV1	Q6alv1 desulfotale
43	72	10.4	475	2	Q8XLZ4	Q8xlz4 clostridium
44	72	10.4	788	2	Q7RJY2	Q7rjy2 plasmodium
45	71.5	10.3	227	2	Q9K884	Q9k884 bacillus ha

ALIGNMENTS

RESULT 1
Q9GZY1
ID Q9GZY1 PRELIMINARY; PRT; 135 AA.
AC Q9GZY1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RA Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1; -.
DR EMBL; AF189269; AAG17117.1; -.
DR Genew; HGNC:21079; PBOV1.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B1983705D0 CRC64;

Query Match 100.0%; Score 694; DB 2; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.9e-60;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAFLRNQKYEDMHNIIHILQIRKLRHRLSNFPRLPGILAPETVLLPFCYKVFRKKEKVK 60

Qy         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTLQLTQTLGLE 120

Qy        121 CCLLYLSKTIHPQII 135
              |||||||||||||||
Db        121 CCLLYLSKTIHPQII 135
```

RESULT 2
O59512
ID O59512 PRELIMINARY; PRT; 381 AA.
AC O59512;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PH1844.
GN OrderedLocusNames=PH1844;
OS Pyrococcus horikoshii.

```
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000007; BAA30965.1; -.
DR   PIR; F71196; F71196.
DR   GO; GO:0009058; P:biosynthesis; IEA.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   381 AA;  42826 MW;  ED0EB0CA0CF5CAE1 CRC64;

  Query Match             13.0%;  Score 90;  DB 2;  Length 381;
  Best Local Similarity   26.8%;  Pred. No. 1.8;
  Matches   34;  Conservative   23;  Mismatches   38;  Indels   32;  Gaps    7;

Qy         13 MHNIIHILQIRKLRHRL----SNFP---------------RLPGILAPETVLLPFCYKVF 53
              |||:   :::|:  | :    :| |                ::|||::|       :
Db         24 MHNL--AIKLRERGHEVGIVTNNRPTGKEEELKRYGIELIKIPGIISP--------FLDV 73

Qy         54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLQTHLTMKGSSM-KCSSLSSEAILFTLTL 111
                :| |::  ||: |: |  |||| |||       | :| | : |:: :| |
Db         74 NLTYGLKSSEELNEFLKDFDIIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132

Qy        112 QLTQTLG 118
              :|  |||
Db        133 KLWDTLG 139


RESULT 3
P93825
ID   P93825      PRELIMINARY;      PRT;   645 AA.
AC   P93825;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   F19P19.25 protein (Putative ethylene receptor).
GN   Name=F19P19.25; Synonyms=ERS2;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Osborne B.I., Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G.,
RA   Buehler E., Conway A.B., Conway A.R., Dewar K., Feng J., Kim C.,
RA   Kurtz D., Li Y., Shinn P., Sun H., Davis R.W., Ecker J.R.,
RA   Federspiel N.A., Theologis A.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Theologis;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Hua J., Sakai1 H., Nourizadeh S., Chen Q.G., Bleecker A.B.,
RA   Ecker J.R., Meyerowitz E.M.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Sakai H., Hua J., Chen Q.G., Chang C., Medrano L.J., Bleecker A.B.,
RA   Meyerowitz E.M.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC000104; AAB70445.1; -.
DR   EMBL; AF047976; AAC62209.1; -.
DR   PIR; F86174; F86174.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR   GO; GO:0007165; P:signal transduction; IEA.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR011052; Prot_amyl_inhib.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00388; HisKA; 1.
KW   Receptor.
SQ   SEQUENCE   645 AA;  72192 MW;  D37ABE3A0939D6ED CRC64;

  Query Match             11.9%;  Score 82.5;  DB 2;  Length 645;
  Best Local Similarity   23.5%;  Pred. No. 17;
  Matches   31;  Conservative   26;  Mismatches   46;  Indels   29;  Gaps    5;

Qy         25 LRHRLSNF---PRLPGILAPETV-----------------LLPFCYKVFRKKEKVKRSQ 63
              : | |: |   |  | ::   ||                 |||   |   |  :   |:
Db         99 MTHLLAGFTYGPHWPWVMTAVTVFKMLTGIVSFLTALSLVTLLPLLLKA--KVREFMLSK 156

Qy         64 KATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKC-SSLSSEAILFTLTLQLTQTLGLECC 122
              | |     :::    |:   :| | ::  : |  :||    ||:|  ::|::|||: |
Db        157 KTRE-----LDREVGIIMKQTETSLHVRMLTTKIRTSLDRHTILYTTLVELSKTLGLKNC 211

Qy        123 LLYLSKTIHPQI 134
               :::   |  ::
Db        212 AVWIPNEIKTEM 223


RESULT 4
Q7SZX8
ID   Q7SZX8      PRELIMINARY;      PRT;   451 AA.
AC   Q7SZX8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   SI:zK13A21.9 (Novel protein similar to vertebrate microphthalmia-
DE   associated transcription factor (MITF) and zebrafish transcription
DE   factor binding to IGHM enhancer 3a (Tfe3a)).
GN   Name=opn1lw2; Synonyms=SI:zK13A21.9;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pandian R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
DR   EMBL; AL844847; CAE30419.1; -.
DR   HSSP; P22415; 1AN4.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
```